BAYESIAN FORECASTING AND DYNAMIC MODELS APPLIED TO STRAIN DATA FROM THE GÖTA RIVER BRIDGE

Ida Kjersem Solhjell

Thesis for the Degree of

MASTER OF SCIENCE

Master in Modelling and Data Analysis

Statistics Division, Department of Mathematics
Faculty of Mathematics and Natural Sciences
University of Oslo

May 2009

Acknowledgments

The writing of this thesis took place in the period from January 2008 to May 2009. During the process I have received a lot of help and support from many people.

In particular, I'm sincerely grateful to my supervisor Bent Natvig, for all his time and thorough guidance. His knowledge, careful editing and prompt feedback have been invaluable. I am also indebted to my co-supervisor Zenon Medina-Cetina for his critical evaluation of my work, and for challenging my ideas. Thanks to NGI for contributing with strain data, and to SMHI for meteorological data.

A special thanks to my fellow students and friends from B800 for many enjoyable moments and good memories.

Finally, but not least, thanks to my family for supporting me in my choices, and for always being there for me. Most of all I want to thank my dear Ivar, for his patience, love, and encouragement.

Ida Kjersem Solhjell
Oslo, May 2009

Contents

CHAPTER 1

Introduction

Systems that change over time have been studied in many scientific fields, and much of the earlier development within time series originated from astronomy and the effort to explain the motion of "heavenly bodies". On January 1st 1801, the Italian astronomer Giuseppe Piazzi, observed the dwarf planet Ceres over 24 nights (corresponding to 9° of its' full orbit), before it disappeared from sight. Based on these imperfect observations, Carl Friederich Gauss was able to estimate the planet's orbit with remarkably high accuracy. In December the same year, Ceres was rediscovered at the exact location Gauss had predicted (Tent (2006)).

Early contributions to time series analysis have mainly been non-Bayesian, while Bayesian forecasting and dynamic models originated in the late 1950s. However, the acclaimed Kalman filter, a recursive procedure for processing data sequentially, has been traced back to 1880 and the Danish statistician and astronomer T.N. Thiele, and is indeed a Bayesian estimator. The developments of Bayesian time series and forecasting has grown enormously in the late 20th century, due to both computational advances and problems related to non-Gaussian, non-linear and non-stationary time series. (Spall (1988),West (1997)).

A time series is a sequence of data assigned to specific moments in time. Most statistical models are static, such as regression analysis: The defining set of parameters has fixed values, and the relationship between the explanatory variables and the response is viewed as constant. This is a perfectly valid assumption for many applications, but when working with time series data, it is important to acknowledge that such relationships may be altered through the passage of time. As opposed to classical time series models, which are static, the Bayesian approach is based on dynamic learning, and allows for varying parameters: As new information is available sequentially, beliefs regarding the parameters expressed through a probability distribution, are updated using Bayes' theorem. Intuitively, recent data are more valuable than older data when making inference on current events. This information loss is recognized when using dynamic models, while for classical time series, all information is weighted equally as the model parameters are static. In addition to its appealing dynamic properties, inference and interpretation of Bayesian time series results are intuitive and straightforward (as for Bayesian statistics in general). As the complete Bayesian time series framework is based on one single theorem, Bayes' theorem, the theory is simplified and unified. The Bayesian paradigm is also particularly suitable for prediction, taking into account all parameter uncertainties, as well as model uncertainty.

In this thesis our main aim is to apply Bayesian dynamic models, as thoroughly presented in West & Harrison (1997), to a univariate strain time series from the Göta River Bridge, and find a model that provides good short and long term predictions. The Göta River Bridge connects Gothenburg's mainland to the island Hissingen. The bridge is over 70 years old, and the steel beams are of relatively poor and varying quality. During the 90s, several minor cracks and fatigue damages were discovered in the bridge structure. The bridge went through major repair, but cracks due to fatigue may occur again, and lead to collapse of the steel girders. Swedish traffic authorities have decided to keep the bridge in service for another 10 to 15 years, but in order to increase safety, the condition of the bridge must be monitored continuously. The Norwegian research center NGI was assigned the job of providing a surveillance system, and has installed over 5km fiber optics to monitor for increasing deformations. The system provides real-time strain data every other hour, for over 50000 points along the bridge girders.

The main framework and theory regarding Baysian time series and dynamic modeling are treated in Chapter 2. Concepts as sequential updating, observation variance learning, use of discount techniques in forecasting and component modeling are introduced. In Chapter 3 the fiber optic monitoring on the Göta River Bridge is described in closer detail, and some basic principles of strain and deformation are reviewed. Both the strain data, and meteorological data from the same period are presented. The Bayesian time series approach to handle missing values is presented, as there are several missing values in the strain dataset. In Chapter 4 the software used in this thesis is discussed, and compared to BATS, a Bayesian time series analysis software provided by Pole et al. (1994). Criteria for evaluating model performance and finding optimal discount factors are introduced in Chapter 5, and several block structured models with trend, season and various types of regression blocks, including transfer functions, are explored. Problems related to forecasting with unknown future regressors are addressed, and solutions suggested. In Chapter 6 a selection of simpler and more complex models are further assessed, and their short, medium and long term prediction is examined throughout different parts of the dataset. Based on the results in Chapter 5 and 6, one model is singeled out, and the model is validated in Chapter 7. Autocorrelations and normal assumption are inspected, and different solutions to correlated errors are presented. In Chapter 8, the principles of "Management by Exception" is introduced, and further a level-shift monitoring scheme with automatic interventions based on Bayes factor is explored. Some of the automatic intervention modes are questioned on basis of unsatisfactory results applying the scheme to the strain data. Finally, in Chapter 9, a summary with some concluding remarks is given, and some topics for further research and model improvement are suggested.

CHAPTER 2

Bayesian Time Series

In this chapter Bayes' theorem, which the complete Bayesian time series analysis is based on, is briefly presented. The ideas of dynamic modeling is introduced, as well as the Bayesian Time series framework and theory, as described in West & Harrison (1997). Concepts as sequential updating, discount strategies and variance learning are treated, and some illustrating examples are given.

2.1. A Short Introduction to Bayes' Theorem

Let the parameter $\boldsymbol{\theta}$ be an unknown quantity. The parameter does have an actual value, but the value is unknown to the scientist, and thus a stochastic value. The *prior distribution* $p(\boldsymbol{\theta})$ states the initial uncertainty concerning the parameter. The prior is constructed by analyzing and quantifying historical information, knowledge and qualified beliefs. To get more information on the parameter, the scientist collects data $D = (x_1, x_2, ..., x_n)$. The joint density of $(x_1, x_2, ..., x_n)$ as a function of $\boldsymbol{\theta}$, is called the *likelihood function* and denoted $p(D|\boldsymbol{\theta})$. To calculate the *posterior distribution*, the initial uncertainty of $\boldsymbol{\theta}$ is updated in light of the new information. The posterior can be found applying the well known Bayes' theorem:

$$p(\boldsymbol{\theta}|D) = \frac{p(\boldsymbol{\theta})p(D|\boldsymbol{\theta})}{p(D)} = \frac{p(\boldsymbol{\theta})p(D|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(D|\boldsymbol{\theta})d\boldsymbol{\theta}} \propto p(\boldsymbol{\theta})p(D|\boldsymbol{\theta})$$

or equivalently Posterior $= \frac{\text{Prior}\times\text{Likelihood}}{\text{Constant}} \propto$ Prior $\times$ Likelihood.

2.2. The Dynamic Linear Model

When working with real time data, the information flow is dynamic. Since Bayesian statistics provides an elegant solution for updating uncertainty, it is well suited for a dynamic context. Bayesian statistics is also excellent for forecasting (as well as decision making), which in many time series applications is the main objective.

Let Y be the response. Then the general linear model is given by $Y = \mathbf{x}'\boldsymbol{\theta} + \nu$, where $\mathbf{x}$ is a vector of known quantities (regressors), $\boldsymbol{\theta}$ is an unknown parameter vector, and ν is a noise term. If the model is time dependent, then

$$Y_t = \mathbf{x}_t'\boldsymbol{\theta}_t + \nu_t$$

This model is static, as Y_t is dependent only on quantities at a given time t. For *dynamic models*, $\boldsymbol{\theta}$ is stochastically dependent over time. The time dependent state parameter is expressed as a first order Markov chain:

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \omega_t$$

where $\mathbf{G}_t$ is a known matrix and ω_t is an error term.

The Dynamic Linear Model(DLM) is introduced in West & Harrison (1997) as a neat and systematic Bayesian approach to handle time series data. The DLM is fully defined by two equations:

$$\begin{aligned} &\textbf{Observation equation:} && Y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \nu_t,\ \nu_t \sim N[0, V_t] \\ &\textbf{System equation:} && \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \omega_t,\ \omega_t \sim N[\mathbf{0}, \mathbf{W}_t] \end{aligned} \tag{2.1}$$

The observation equation defines the relationship between the observed data and the unknown state parameters, and the system equation describes the evolution of the state parameter over time. The notations used throughout this thesis are:

- **Observations:** Y_t denotes the observation series at time t.
- **Regression vector:** $\mathbf{F}_t$ is a $(n \times 1)$ column vector of known regressors.
- **State parameter:** $\boldsymbol{\theta}_t$ denotes the $(n \times 1)$ column vector of unknown parameters.
- **Observation error:** ν_t is a normally distributed stochastic error term with a zero mean.
- **Observation variance:** V_t is the variance of the observation errors.
- **Evolution matrix:** $\mathbf{G}_t$ is an $(n \times n)$ matrix of known coefficients. $\mathbf{G}_t$ defines the relationship between the parameters at time t and time $t-1$, and thus determines the evolution of the state parameters.
- **Evolution error:** ω_t is a stochastic error term which can be said to describe information loss over time. ω_t is normally distributed with zero mean.
- **Evolution covariances:** $\mathbf{W}_t$ is an $(n \times n)$ covariance matrix for the evolution error.

It is assumed that $\mathbf{W}_t$ and V_t are independent. Consider a time series $\{Y_1, Y_2, ...\}$. At time $t = 0$, before observing any data, there might be available information D_0 on the parameter $\boldsymbol{\theta}$, such as expert knowledge or historical information. This is called *initial information*, and can be quantified through a probability distribution

$$\boldsymbol{\theta}_0 | D_0 \sim N[\mathbf{m}_0, \mathbf{C}_0]$$

for a suitable $\mathbf{m}_0$ and $\mathbf{C}_0$.

EXAMPLE 2.1. *Assume that a strain time series can be modeled by the simple observation equation*

$$strain_t = level_t + \beta_t temperature_t + \nu_t, \quad \nu_t \sim N[0, V]$$

Given a known observation variance V, there are two unknown quantities in this equation, namely $level_t$ and β_t. These two parameters are further described as simple random walks:

$$level_t = level_{t-1} + \omega_{1,t}, \quad \omega_{1,t} \sim N[0, W_1]$$

$$\beta_t = \beta_{t-1} + \omega_{t,2}, \quad \omega_{2,t} \sim N[0, W_2]$$

Since ν_t, $\omega_{1,t}$ and $\omega_{2,t}$ have zero means, we expect that $strain_t = strain_{t-1}$, but it may vary due to the uncertainty in the model. The extent of variation in the 1-step predictions are determined by the values V, W_1 and W_2. If $V, W_1, W_2 = 0$, the model is static, and the parameters have a fixed value through time. When V, W_1 or W_2 increase, so does the volatility of the parameters. The model is defined as a dynamic linear model as follows:

$$Y_t = strain_t, \quad \boldsymbol{\theta}_t = (level_t, \beta_t)', \quad \mathbf{F}_t = (1, temperature_t)',$$

$$\mathbf{G}_t = \mathbf{G} = \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

□

2.3. Observation Variance Learning

In real applications V_t is rarely known, and further it is convenient to work with a constant observation variance $V_t = V$. Observation errors are typically related to measurement errors, instrumental errors etc. and are in most cases more or less constant. If changes in the observation variance is expected, for instance if new measurement methods are introduced, subjective intervention can be made to allow a shift in V. Being true to the Bayesian approach, all beliefs and uncertainties concerning V are quantified through a proper prior probability distribution, and updated whenever new information is available. It is practical to work with V indirectly through the precision, $\phi = 1/V$. The initial information on the precision is expressed as a gamma distribution:

$$(\phi|D_0) \sim \Gamma(n_0/2, S_0 n_0/2), \tag{2.2}$$

where n_0 is some known constant and the expected value of ϕ is S_0^{-1}, thus S_0 is a prior point estimate for V. Since the gamma distribution is a natural conjugate prior for the T-distribution, the updating is neat and tidy. Even though V is assumed to be constant, the updating scheme allows for small, slow variations in the observation variance. The DLM with unknown observation variance is defined by

$$\begin{aligned} &\textbf{Observation equation:} && Y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \nu_t, \ \nu_t \sim N[0, V] \\ &\textbf{System equation:} && \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \omega_t, \ \omega_t \sim T_{n_{t-1}}[\mathbf{0}, \mathbf{W}_t] \end{aligned}$$

where n_{t-1} denotes the degrees of freedom. The initial information is given by:

$$\boldsymbol{\theta}_0|D_0 \sim T_{n_0}[\mathbf{m}_0, \mathbf{C}_0],$$

where $\mathbf{m}_0$ is a prior point estimate for $\boldsymbol{\theta}_0$ and $\mathbf{C}_0$ reflects the uncertainty of the parameter.

2.4. Sequential Updating

Assume that the *posterior distribution* $p(\boldsymbol{\theta}_t|D_t)$ is known, based on the observed data $\{Y_1, Y_2, ..., Y_t\}$ and any other relevant information. The prior distribution on the state parameters for time $t+1$ is easily constructed by applying the system equation. Given $\mathrm{E}[\boldsymbol{\theta}_t|D_t] = \boldsymbol{m}_t$ and $\mathrm{Var}[\boldsymbol{\theta}_t|D_t] = \mathbf{C}_t$, then

$$\mathrm{E}[\boldsymbol{\theta}_{t+1}|D_t] = \mathbf{a}_{t+1} = \mathbf{G}_{t+1}\,\mathrm{E}[\boldsymbol{\theta}_t|D_t] + \mathrm{E}[\omega_t] = \mathbf{G}_{t+1}\mathbf{m}_t$$

$$\mathrm{Var}[\boldsymbol{\theta}_{t+1}|D_t] = \mathbf{R}_{t+1} = \mathbf{G}_{t+1}\,\mathrm{Var}[\boldsymbol{\theta}_t|D_t]\mathbf{G}'_{t+1} + \mathrm{Var}[\omega_{t+1}] = \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}'_{t+1} + \mathbf{W}_{t+1}$$

The state parameter prior at time $t+1$ is simply an adjustment of the posterior at time t. The parameters are shifted due to the evolution process given by $\mathbf{G}_{t+1}$. W_{t+1} represents the extra uncertainty added by predicting the state parameters one step ahead.

The *1-step forecast distribution*, $p(Y_{t+1}|D_t)$, is found applying the observation equation:

$$\mathrm{E}[Y_{t+1}|D_t] = f_{t+1} = \mathbf{F}'_{t+1}E[\boldsymbol{\theta}_{t+1}|D_{t+1}] + E[\nu_{t+1}] = \mathbf{F}'_{t+1}\mathbf{a}_{t+1}$$

$$\mathrm{Var}[Y_{t+1}|D_t] = Q_{t+1} = \mathbf{F}'_{t+1}\,\mathrm{Var}[\boldsymbol{\theta}_{t+1}|D_{t+1}]\mathbf{F}_{t+1} + \mathrm{Var}[\nu_t] = \mathbf{F}'_{t+1}\mathbf{a}_{t+1}\mathbf{F}_{t+1} + V,$$

where S_t is used as a point estimate for V if the observation variance is unknown.

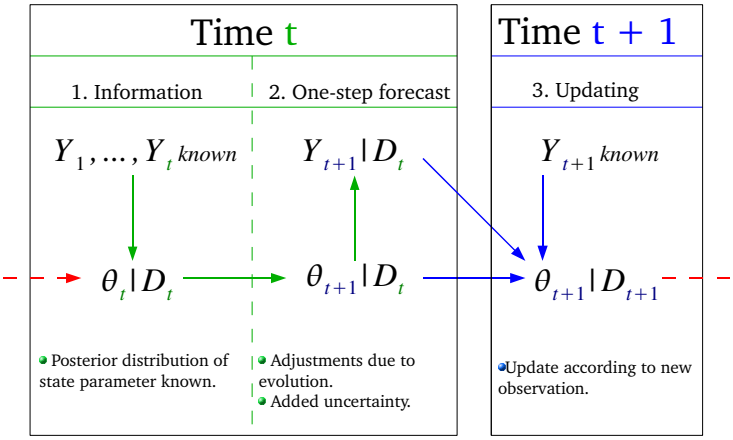


Figure 2.1: DLM updating

When Y_{t+1} is available, the posterior distribution for time $t+1$, $p(\boldsymbol{\theta}_{t+1}|D_{t+1})$, can be calculated by applying Bayes' theorem:

$$p(\boldsymbol{\theta}_{t+1}|D_{t+1}) \propto p(\boldsymbol{\theta}_{t+1}|D_t)p(Y_{t+1}|\boldsymbol{\theta}_{t+1}),$$

where $D_{t+1} = \{D_t, Y_{t+1}\}$. Figure 2.1 illustrates the DLM dynamics.

A univariate DLM with a constant unknown variance $V = \phi^{-1}$ gives the following updating relationships:

State posterior at time t:	$(\boldsymbol{\theta}_t \mid D_t) \sim T_{n_t}[\mathbf{m}_t, \mathbf{C}_t]$
State prior at time t:	$(\boldsymbol{\theta}_{t+1} \mid D_t) \sim T_{n_t}[\mathbf{a}_{t+1}, \mathbf{R}_{t+1}]$
Precision information at time t:	$(\phi \mid D_t) \sim G[\frac{n_t}{2}, \frac{n_t S_t}{2}]$
	$\mathbf{a}_{t+1} = \mathbf{G}_{t+1}\mathbf{m}_t$
	$\mathbf{R}_{t+1} = \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}'_{t+1} + \mathbf{W}_{t+1}$
1-step forecast distribution at time t:	$(Y_{t+1} \mid D_t) \sim T_{n_t}[f_{t+1}, Q_{t+1}]$
	$f_{t+1} = \mathbf{F}'_{t+1}\mathbf{a}_{t+1}$
	$Q_{t+1} = \mathbf{F}'_{t+1}\mathbf{R}_{t+1}\mathbf{F}_{t+1} + S_t$
State posterior at time $t+1$:	$(\boldsymbol{\theta}_{t+1} \mid D_{t+1}) \sim T_{n_{t+1}}[\mathbf{m}_{t+1}, \mathbf{C}_{t+1}]$
Precision information at time $t+1$:	$(\phi \mid D_{t+1}) \sim G[\frac{n_{t+1}}{2}, \frac{n_{t+1}S_{t+1}}{2}]$
	$\mathbf{m}_{t+1} = \mathbf{a}_{t+1} + \mathbf{A}_{t+1}e_{t+1}$
	$\mathbf{C}_{t+1} = \frac{S_{t+1}}{S_t}[\mathbf{R}_{t+1} - \mathbf{A}_{t+1}\mathbf{A}'_{t+1}Q_{t+1}]$
	$e_{t+1} = Y_{t+1} - f_{t+1}$
	$\mathbf{A}_{t+1} = \frac{\mathbf{R}_{t+1}\mathbf{F}_{t+1}}{Q_{t+1}}$
	$n_{t+1} = n_t + 1$
	$S_{t+1} = S_t + \frac{S_t}{n_{t+1}}\left(\frac{e^2_{t+1}}{Q_{t+1}} - 1\right)$

For further reading and proofs, see Theorem 4.3 in West & Harrison (1997). If ν_t is non-Gaussian, the focus is on moments rather than fully specified distributions.

2.5. Variance Discounting

As for V, the evolution covariance matrix $\mathbf{W}_t$ is typically unknown. In Section 2.3 a learning procedure for V was introduced. The same can be done for W_t, but when the state parameter vector $\boldsymbol{\theta}_t$ is multidimensional, the implementation is complex. A reasonable and functional alternative is information discounting. A discount factor δ is a measure of information loss through the evolution process. Pole et al. (1994) operates with information loss being equal to $1 - \delta$. If the information loss from time t to time $t+1$ is 5%, then $\delta = 0.95$. Earlier we defined

$$\mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t + \mathbf{W}_t$$

With variance discounting

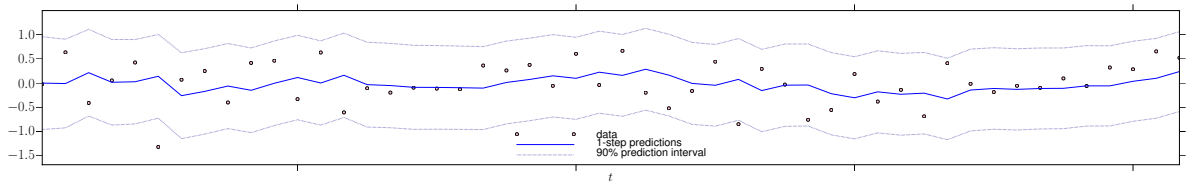
$$\mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \mathbf{R}_t = \delta^{-1}\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t, \quad \delta \in [0, 1]$$

This further implies that

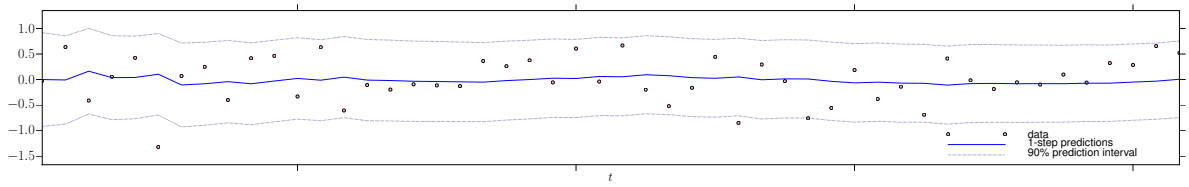
$$\mathbf{W}_t = \frac{1-\delta}{\delta}\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t',$$

thus $\mathbf{W}_t$ is invariant to measurement scale. As $\mathbf{W}_t$ is inversely proportional to δ, low discount factors will result in much added uncertainty when predicting. High prediction uncertainty indicates lack of knowledge on the parameters. When updating to a posterior distribution, both the prior distribution and the new observation is weighted. If the prior is very uncertain, the new observation is weighted heavily. As a consequence new data will be emphasized more than earlier data when updating, if the discount factor is low. Hence the parameter vector will as a consequence have a short term memory, indicating large information loss. A low discount factor allows the parameters to adjust quickly. If the discount factor is too low, the estimated time series will be volatile as a consequence of modeling noise, and it will be sensitive to outliers. Regardless the amount of information, the parameters will only depend on recent observations. A high discount factor gives long term memory with stable parameters which do not respond significantly to noise or outliers. These are certainly good qualities, but if the discount factor is chosen too high, the parameters will not adjust to actual changes. The optimal discount factor compromises between stability and flexibility. How to choose the optimum is reviewed in Chapter 5.

EXAMPLE 2.2. *In the following synthetic case, the data* $\{Y_1, Y_2, ..., Y_{50}\}$ *are normally distributed with a zero mean and* 0.2 *variance. The obvious choice of model is a constant trend model with level as the single parameter, also called a 1st order polynomial trend. Two constant trend models with different discount factors are constructed. The first model has a 0.75 discount factor, implying 25% information loss, while the second model uses a 0.95 discount factor, equivalent to 5% information loss. From Figure 2.2, it is obvious that the latter model is best suited for the data. Figure 2.2 (a) shows that by using* $\delta = 0.75$*, the predictions are heavily influenced by noise. For the second model with* $\delta = 0.95$*, the predictions shown in Figure 2.2 (b) are stable and in correspondence with the real underlying level. For more irregular data, a lower discount factor might do better, but in general discount factors less than 0.8 are considered too small as the available information is not fully utilized.* □



(a) $\delta = 0.75$



(b) $\delta = 0.95$

Figure 2.2: *1-step predictions with different discount factors.*

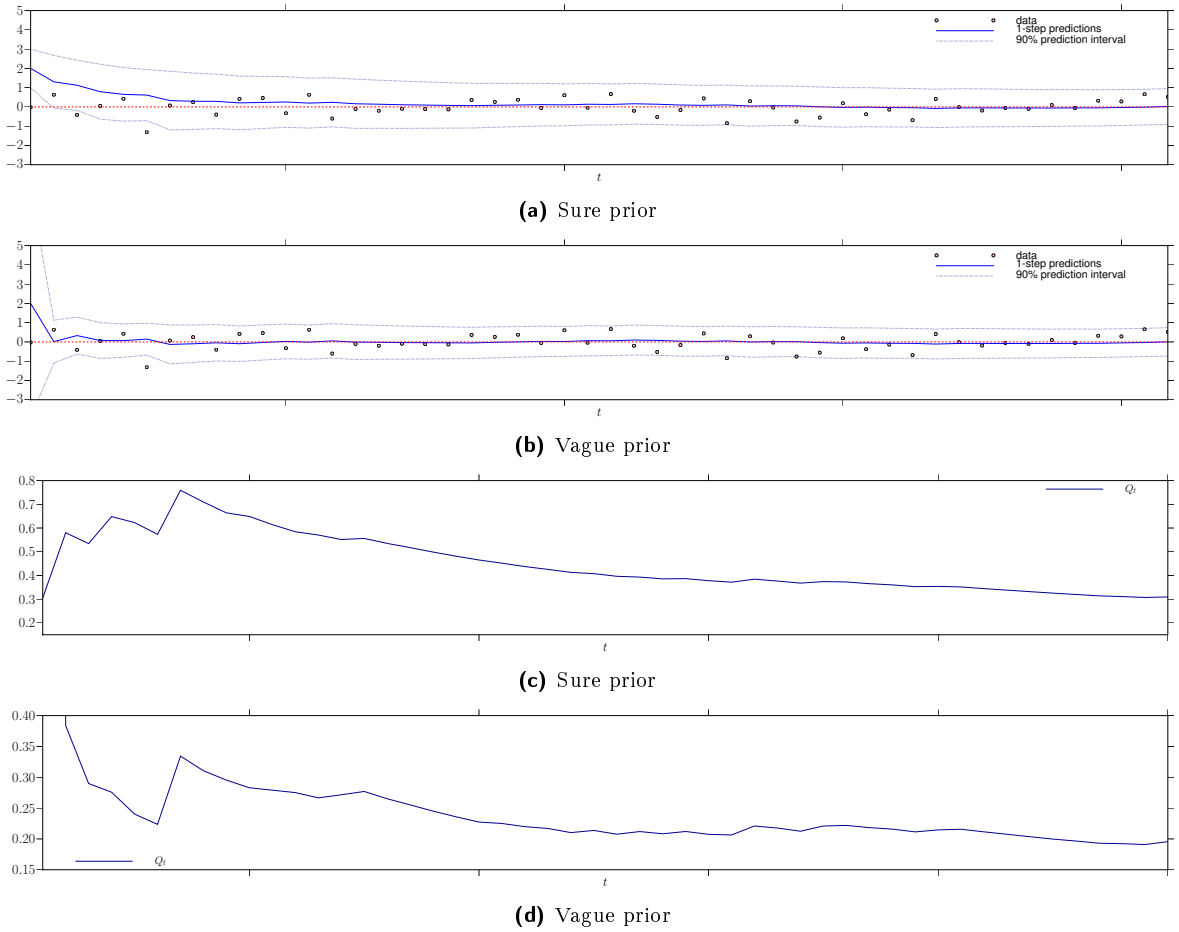
2.6. Quantifying Initial Information

Initial information is the probabilistic representation of the forecaster's beliefs about the parameters, where the prior point estimates for the state parameters and the corresponding uncertainty are set by $\mathbf{m}_0$ and $\mathbf{C}_0$ respectively. Little or no initial knowledge on the parameters is reflected through a vague prior. By setting the initial variances on the diagonal of $\mathbf{C}_0$ adequately large, the choice of $\mathbf{m}_0$ is not crucial, as the state parameters will quickly adjust to the observed data. The remaining values of the variance matrix $\mathbf{C}_0$ are typically set to 0. The following example will illustrate how state parameters adjust to data using vague and sure priors (the latter indicating certain knowledge).

EXAMPLE 2.3. *Two constant trend models are constructed to analyze the same synthetic data as in Example 2.2. The initial expected level m_0 is set to 2 for both models. The real underlying level is 0, so the initial belief is incorrect. The first model has a sure prior, with variance $C_0 = 0.1$, indicating that the initial guess is well justified. For the second model $C_0 = 10$, indicating uncertain knowledge through a vague prior. From Figures 2.3 (a) and (b), which shows 1-step predictions using the sure and the vague prior respectively, it is evident that the level parameters reaches the correct level much more efficiently using a vague prior. It is also clear by Figures 2.3 (c) and (d) that the latter ensures more rapid convergence of the prediction variance. The sure prior would obviously be the better choice if the m_0 hit target.*

In this example the trend discount factor value used was 0.95, indicating a 5% information loss for each step of time. A high discount factor is a natural choice for such stable data. If a lower value had been used, the level of the sure prior would be able to adjust faster, but when the parameter reaches the real level, the model would be too sensitive to noise.

If there are enough data available, a possibility is to use a portion of the data as a training set. The training set is then analyzed using a vague prior. The parameter values reached at the end



(a) Sure prior

(b) Vague prior

(c) Sure prior

(d) Vague prior

Figure 2.3: *(a),(b): 1 step predictions. (c),(d):Prediction variances*

of the set can further be used as a prior for the remaining part of the dataset. This way it is ensured that a potentially volatile burn in period is omitted when comparing model performances. □

2.7. k-step Predictions

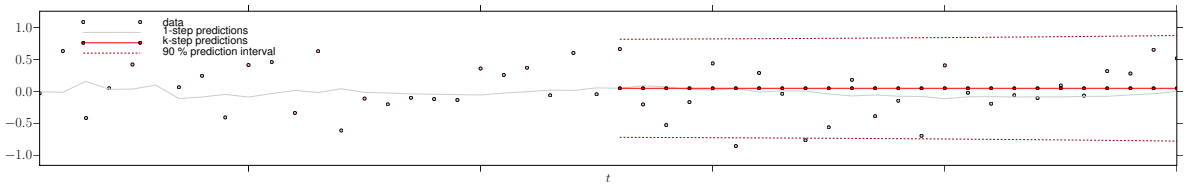
k-step predictions are calculated by projecting the prior several steps into the future. Given the posterior distribution $p(\boldsymbol{\theta}_t|D_t)$ for the state parameter at time t, the system equation is applied k times to get $p(\boldsymbol{\theta}_{t+k}|D_t)$, the distribution of the state parameter at time $t+k$ given all available information at time t. To calculate the desired forecast distribution $p(Y_{t+k}|D_t)$, the observation equation is applied to the obtained $\boldsymbol{\theta}_{t+k}$. Predictions are updated whenever new information is available. For a DLM with unknown, constant observation variance the following relationships

apply (from summary of Theorem 4.2, page 112 in West & Harrison (1997)):

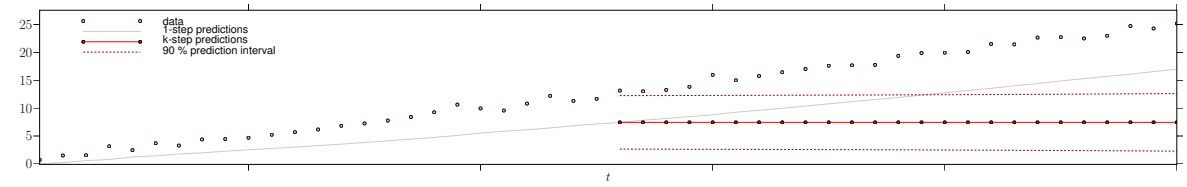
k-step forecast:	$(\boldsymbol{\theta}_{t+k}\|D_t) \sim T_{n_t}[\mathbf{a}_t(k), \mathbf{R}_t(k)]$, for $k \geq 1$
	$(Y_{t+k}\|D_t) \sim T_{n_t}[f_t(k), Q_t(k)]$, for $k \geq 1$
	$\mathbf{a}_t(k) = \mathbf{G}_{t+k}\mathbf{a}_t(k-1)$
	$\mathbf{R}_t(k) = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}'_{t+k} + \mathbf{W}_{t+k}$
	$f_t(k) = \mathbf{F}'_{t+k}\mathbf{a}_t(k)$
	$Q_t(k) = \mathbf{F}'_{t+k}\mathbf{R}_t(k)\mathbf{F}_{t+k} + S_t$
	$\mathbf{a}_t(0) = \mathbf{m}_t$
	$\mathbf{R}_t(0) = \mathbf{C}_t$
State covariances:	$C[\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_{t+j}\|D_t] = \mathbf{C}_t(k,j)$
Observation covariances:	$C[Y_{t+k}, Y_{t+j}\|D_t] = \mathbf{F}'_{t+k}\mathbf{C}_t(k,j)\mathbf{F}_{t+j}$
Other covariances:	$C[\boldsymbol{\theta}_{t+k}, Y_{t+j}\|D_t] = \mathbf{C}_t(k,j)\mathbf{F}_{t+j}$
	$C[Y_{t+k}, \boldsymbol{\theta}_{t+j}\|D_t] = \mathbf{F}'_{t+k}\mathbf{C}_t(k,j)$
	$\mathbf{C}_t(k,j) = \mathbf{G}_{t+k}\mathbf{C}_t(k-1,j),\ k = j+1, ...$
	$\mathbf{C}_t(j,j) = \mathbf{R}_t(j)$

For $k = 1$ $f_t(1) = f_t$ and $Q_t(1) = Q_t$. For each step k the term $\mathbf{W}_{t+k}$ is added to the prediction variance to represent the extra uncertainty induced by predicting into the future.

EXAMPLE 2.4. *In this example the same model is applied to two different synthetic datasets. The first dataset has an underlying constant trend, while the second has an underlying linear trend. Both datasets are analyzed with a constant trend model. At time $t = 25$, predictions are made 25 steps ahead. The data in Figure 2.4 (a) have an underlying constant trend, and a constant trend model is obviously the correct choice, and gives accurate step ahead predictions. The data in Figure 2.4 (b) have an underlying linear trend, and thus the constant trend model is inadequate, since it only has a level parameter, and does not model the slope. The level parameter adjusts every time new data are observed, but the step ahead predictions are only as complex as the input model, and since a constant trend model is used, the future dynamics of the time series is expected to be constant.* □



(a) Constant underlying trend



(b) Linear underlying trend

Figure 2.4: *k-step predictions with 90 % prediction interval.*

2.8. Retrospective Analysis

The aim of retrospective analysis is to estimate what has happened in the past, in light of all current information. For a specific point of time the retrospective analysis is based both on information previous and subsequent to this event. To analyze data in retrospect, also called smoothing or filtering, can give an integrated vision of the time series, so the nature of the series is better understood. It is often easier to analyze what has happened on a given point of time, when seen in a larger context. The following definitions are given in Corollary 4.3 and Theorem 4.5 West & Harrison (1997):

Retrospective state distribution: $(\boldsymbol{\theta}_{t-k}|D_t) \sim T_{n_t}[a_t(-k), (S_t/S_{t-k})\mathbf{R}_t(-k)], \quad \text{for } k \geq 0$

Retrospective state covariances: $C[\boldsymbol{\theta}_{t-k-j}, \boldsymbol{\theta}_{t-k}|D_t] = A_{t-k-j,t-k}\mathbf{R}_t(-k), \quad \text{for } j, k \geq 0$

$$\mathbf{a}_t(-k) = \mathbf{m}_{t-k} - \mathbf{B}_{t-k}[\mathbf{a}_{t-k+1} - \mathbf{a}_t(-k+1)]$$
$$\mathbf{R}_t(-k) = \mathbf{C}_{t-k} - \mathbf{B}_{t-k}[\mathbf{R}_{t-k+1} - \mathbf{R}_t(-k+1)]\mathbf{B}'_{t-k}$$
$$\mathbf{B}_t = \mathbf{C}_t\mathbf{G}'_{t+1}\mathbf{R}^{-1}_{t+1}$$
$$\mathbf{a}_t(0) = \mathbf{m}_t$$
$$\mathbf{R}_t(0) = \mathbf{C}_t$$

Retrospective analysis is also known as smoothing or filtering.

2.9. Component modeling

Applying the practice described in West & Harrison (1997) three main component forms are used to recognize the behavior of a complex time series:

(1) *Simple trends.* This component accounts for the evolution of the underlying trend in the time series. It is a standard component in any model, unless the level has a fixed zero value.

(2) *Systematic cyclical variation.* Cyclic variation is typically related to seasons, time of day, etc. To model cyclical behavior, an alternative is to introduce seasonal effects, a vector of constants which represents seasonal departures from the trend. It is required that this vector sums up to zero. A more practical approach is to use linear combinations of sine and cosine functions, designed to provide the desired period and amplitude of the periodic pattern.

(3) *Influential causal variables.* A regression component models any influence of explanatory variables.

A DLM combining these three components provides a flexible tool to describe simple or advanced time series dynamics by the *superpositioned* DLM

$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu \sim N[0, V],$$

equivalent to Data = Trend + Seasonal variation + Regression + Error. The superpositioned DLM, or block DLM, is then defined by

$$\begin{gathered} \boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{T_t} \\ \boldsymbol{\theta}_{S_t} \\ \boldsymbol{\theta}_{R_t} \end{pmatrix}, \qquad \mathbf{F}_t = \begin{pmatrix} \mathbf{F}_{T_t} \\ \mathbf{F}_{S_t} \\ \mathbf{F}_{R_t} \end{pmatrix}, \\ \mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{T_t} & 0 & 0 \\ 0 & \mathbf{G}_{S_t} & 0 \\ 0 & 0 & \mathbf{G}_{R_t} \end{pmatrix}, \qquad \mathbf{W}_t = \begin{pmatrix} \mathbf{W}_{T_t} & 0 & 0 \\ 0 & \mathbf{W}_{S_t} & 0 \\ 0 & 0 & \mathbf{W}_{R_t} \end{pmatrix}, \end{gathered} \tag{2.3}$$

where T is the trend block, S is the season block and R the regression block.

2.9.1. Trend Block. Low order polynomial functions are typically used to describe smooth developments over time. A *1st order polynomial trend* gives a constant level model, and the following trend DLM block:

$$\begin{aligned} Y_{T_t} &= \alpha_t \\ \alpha_t &= \alpha_{t-1} + \omega_t, \quad \omega_t \sim N[0, W], \end{aligned}$$

where α_t corresponds to the level.

A *2nd order polynomial trend* provides a linear growth model. The trend DLM block is then

$$\begin{aligned} Y_{T_t} &= \alpha_t \\ \alpha_t &= \alpha_{t-1} + \mu_{t-1} + \omega_{t,1}, \quad \omega_{t,1} \sim N[0, W_1] \\ \mu_t &= \mu_{t-1} + \omega_{t,2}, \quad \omega_{t,2} \sim N[0, W_2] \end{aligned}$$

$$\begin{aligned} \boldsymbol{\theta}_t &= (\alpha_t, \mu_t)', \quad \mathbf{F}_t = (1, 0) \\ \mathbf{G}_t &= \mathbf{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \end{aligned}$$

where μ_t represents the change of rate. 1st and 2nd order polynomial trends are most common, and orders higher than 3 are rarely used.

2.9.2. Seasonal Block. The season block models periodical behavior, typically related to time of year, time of day etc. One of several ways to describe cyclical behavior is to use seasonal effects, representing deviations around the trend. The sum of all effects during one period must sum up to zero. Let p denote the period. Then the a DLM block with seasonal effects is defined as follows:

$$\begin{aligned} \boldsymbol{\theta}_S &= \phi = (\phi_1, ..., \phi_p), \quad \mathbf{F}_{S_t} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}' \\ \mathbf{G}_{S_t} &= \mathbf{G}_S = \begin{pmatrix} 0 & \mathbf{I}_{p-1} \\ 1 & \mathbf{0} \end{pmatrix} \end{aligned}$$

Thus if $\boldsymbol{\theta}_{S_t} = \phi_1$ then $\boldsymbol{\theta}_{S_{t+1}} = \phi_2$ and so forth. An alternative to form-free effects are Fourier representations, where each effect corresponds to a combination of sine and cosine functions. A cyclical pattern with period p, can be represented by a cosine function $a_t \cos(\omega(t-1))$, where $\omega = \frac{2\pi}{p}$ and a_t is the amplitude. The function obviously has a maximum when $t = 1$, and thus represents the season peak. If the season peak is located elsewhere, it can easily be shifted by adding a sine term: $a_t \cos(\omega(t-1)) + b_t \sin(\omega(t-1))$. The function is called a *single harmonic* and can be expressed as a DLM block:

$$\begin{aligned} \boldsymbol{\theta}_{S_t} &= (a_t, b_t)' \\ \mathbf{F}_{S_t} &= \mathbf{F}_S = (1, 0)' \\ \mathbf{G}_{S_t} &= \mathbf{G}_S = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \end{aligned}$$

Increasing the number of harmonics, equally increases the complexity of the seasonal structure. The number of harmonics corresponds to the number of peaks during one cycle. A DLM with n

harmonics can be written as

$$\begin{aligned}\boldsymbol{\theta}_{S_t} &= (a_{t,1}, ..., a_{t,n}, b_{t,1}, ..., b_{t,n})' \\ \mathbf{F}_{S_t} &= \mathbf{F}_S = (1, 0, 1, ..., 0)' \\ \mathbf{G}_{S_t} &= \mathbf{G}_S = \text{diag}(G_{S_1}, ..., G_{S_n}) \\ \mathbf{G}_{S_j} &= \begin{pmatrix} \cos j\omega & \sin j\omega \\ -\sin j\omega & \cos j\omega \end{pmatrix}\end{aligned}$$

Working with harmonic components is convenient, but a seasonal effect at time t can not be directly interpreted from the parameters a_t and b_t. It is therefore practical to calculate the seasonal effects, $\boldsymbol{\phi}_t$, by doing the linear transform $\boldsymbol{\phi}_t = \mathbf{L}\boldsymbol{\theta}_{S_t}$ where

$$\mathbf{L} = \begin{pmatrix} \mathbf{F}'_S \\ \mathbf{F}'_S\mathbf{G}_S \\ \mathbf{F}'_S\mathbf{G}_S^2 \\ \vdots \\ \mathbf{F}'_S\mathbf{G}_S^{p-1} \end{pmatrix} \tag{2.4}$$

2.9.3. Regression Block. A linear regression on some variable X_t can be expressed as a DLM block as follows:

$$\begin{aligned}Y_{\mathbf{R}_t} &= \alpha_t X_t \\ \alpha_t &= \alpha_{t-1} + \omega_t, \qquad \omega_t \sim N[0, W]\end{aligned}$$

$$\boldsymbol{\theta}_t = \alpha_t, \quad \mathbf{F}_t = X_t, \quad \mathbf{G}_t = \mathbf{G} = 1$$

When regressing on several variables $\mathbf{X} = X_1, ..., X_n$, the regression DLM has the following form:

$$\begin{aligned}Y_{\mathbf{R}_t} &= \beta_{1t}X_{1t} + ... + \beta_{nt}X_{nt} + \nu_t, \qquad \nu_t \sim N[0, V] \\ \beta_{it} &= \beta_{it-1} + \omega_{it}, \qquad \omega_{t,i} \sim N[0, W_i], \quad i = 1, ..., n \\ \boldsymbol{\theta}_t &= (\beta_{1t}, ...\beta_{nt})', \quad \mathbf{F}_t = \mathbf{X}_t, \quad \mathbf{G}_t = \mathbf{G} = \mathbf{I}_n\end{aligned}$$

In most application, future regression coefficients are unknown, which complicates the forecasting process. Various solutions to this problem will be presented in Chapter 5.

2.9.4. Component Discounting. For a single component model the prior variance $\mathbf{R}_t$ is defined as:

$$\mathbf{R}_t = \delta^{-1}\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t \tag{2.5}$$

When working with a superpositioned model, the information loss for different components is not necessarily equal. The season block is for example typically more stable than the trend block. One single discount factor is therefore not sufficient for superpositioned models. West &

Harrison (1997) (page 196-198) suggests using separate discount factors for each block. For a model with h components, the discount factors are $\boldsymbol{\delta} = \{\delta_1, ..., \delta_h\}$. Let $\mathbf{P}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t'$. Then the block diagonal elements of $\mathbf{R}_t$ are defined by

$$\{\mathbf{R}_t\}_i = \delta_i^{-1}\{\mathbf{P}_t\}_i$$

where i specifies the block. For a model with only one block, this is in correspondence with a single discount model. For a model with more than one block, the remaining elements of $\mathbf{R}_t$, which corresponds to the prior covariance between parameters from different blocks (e.g. between a parameter from a regression block and from a trend block) are set to

$$\{\mathbf{R}_t\}_{i,j} = \{\mathbf{P}_t\}_{i,j} \tag{2.6}$$

where i and j specifies row and column respectively. In other words these prior covariances are not influenced by any discount factor. Intuitively, the discount factors should not influence the correlation between any two parameters. For a single component DLM this applies as

$$\begin{aligned}
\text{Corr}(\theta_{i,t}, \theta_{j,t}|D_{t-1}) &= \frac{\text{Cov}(\theta_{i,t}, \theta_{j,t}|D_{t-1})}{\sqrt{\text{Var}(\theta_{i,t}|D_{t-1})}\sqrt{\text{Var}(\theta_{j,t}|D_{t-1})}} \\
&= \frac{\delta^{-1}\{\mathbf{P}_t\}_{i,j}}{\sqrt{\delta^{-1}\{\mathbf{P}_t\}_{i,i}}\sqrt{\delta^{-1}\{\mathbf{P}_t\}_{j,j}}} \\
&= \frac{\{\mathbf{P}_t\}_{i,j}}{\sqrt{\{\mathbf{P}_t\}_{i,i}}\sqrt{\{\mathbf{P}_t\}_{j,j}}}
\end{aligned}$$

This seems reasonable, but unfortunately we don't get the same result using component discounting. Let θ_i and θ_j be parameters from two different components c_1 and c_2. Then

$$\begin{aligned}
\text{Corr}(\theta_{i,t}, \theta_{j,t}|D_{t-1}) &= \frac{\text{Cov}(\theta_{i,t}, \theta_{j,t}|D_{t-1})}{\sqrt{\text{Var}(\theta_{i,t}|D_{t-1})}\sqrt{\text{Var}(\theta_{j,t}|D_{t-1})}} \\
&= \frac{\{\mathbf{P}_t\}_{i,j}}{\sqrt{\delta_{c_1}^{-1}\{\mathbf{P}_t\}_{i,i}}\sqrt{\delta_{c_2}^{-1}\{\mathbf{P}_t\}_{j,j}}}
\end{aligned}$$

An alternative component discounting method is briefly presented on page 201-202 in West & Harrison (1997). Let $\Delta = diag(\delta_1^{-1/2}, ..., \delta_1^{-1/2})$. Then the prior variance is defined by

$$\mathbf{R}_t = \Delta\mathbf{P}_t\Delta \tag{2.7}$$

Again let θ_i and θ_j be parameters from two different components c_1 and c_2. Then

$$\begin{aligned}
\text{Corr}(\theta_{i,t}, \theta_{j,t}|D_{t-1}) &= \frac{\text{Cov}(\theta_{i,t}, \theta_{j,t}|D_{t-1})}{\sqrt{\text{Var}(\theta_{i,t}|D_{t-1})}\sqrt{\text{Var}(\theta_{j,t}|D_{t-1})}} \\
&= \frac{\sqrt{\delta_{c_1}^{-1}}\sqrt{\delta_{c_2}^{-1}}\{\mathbf{P}_t\}_{i,j}}{\sqrt{\delta_{c_1}^{-1}\{\mathbf{P}_t\}_{i,i}}\sqrt{\delta_{c_2}^{-1}\{\mathbf{P}_t\}_{j,j}}} \\
&= \frac{\{\mathbf{P}_t\}_{i,j}}{\sqrt{\{\mathbf{P}_t\}_{i,i}}\sqrt{\{\mathbf{P}_t\}_{j,j}}}
\end{aligned}$$

which gives the same satisfying result as for the single component model for all parameter correlations. Obviously, we have that $\mathbf{W}_t = \mathbf{R}_t - \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t'$. Thus using this discounting strategy will usually not lead to a block diagonal evolution noise matrix $\mathbf{W}_t$. We use the latter discount method throughout this thesis.

A discount factor between 0.8 and 1.0 is typically appropriate for the trend component. Lower values might be evidence that the model is insufficient. If for example critical explanatory variables are left out, the trend will compensate for the missing regression contribution, and hence the trend discount factors must be low to allow the trend parameter to be flexible. The season discount factor is typically high, as new information for one specific seasonal effect is only collected every pth observation.

CHAPTER 3

Data

3.1. The Göta River Bridge

The Göta River Bridge was built in 1939 and expanded in 1958. According to the article NGI (2008a), the bridge carries around 26000 vehicles a day, and is an important artery for local traffic, public transport and pedestrians. Due to fatigue damages and several minor cracks, the bridge went through major restoration from 1996 to 1999, with a cost of 110 million SEK. It was later decided that the bridge can be used until 2020, conditioned that the traffic load is to be limited, and the bridge monitored continuously. In 2005 NGI was assigned by the local traffic authorities to install a surveillance system. After evaluating different alternatives, NGI landed on the Brillouin scattering method. The method was chosen because it gives detailed, precise and cost efficient real-time measurements.

The bridge has seven longitudinal steel girders stretching along the length of the bridge. Along five of the girders, more then 5 km of fiber optic cables has been installed to monitor deformations and damages on the bridge, using the Brillouin method. In essence the method involves sending light into one end of a fiber optic cable, and measure the light when it comes out in the other end of the cable, and its reflection. If the cable has been stretched or compressed, the frequency of the light will change. The measurements are very precise, and returns micro strain values for each 10cm of the monitored girders. With over 5 km of monitored fiber optic there are all in all over 50000 monitored points (10cm stretches) on the bridge, so failures can be located with very high accuracy. It takes two hours for the light to pass through one cable loop, so real-time data are reported for the entire bridge every other hour.

Figure 3.1: The Göta River Bridge. (Photo: Stig Hedström, Vägverket)

3.2. The Bridge Coordinate System

The Göta River Coordinate System is a mapping system for the physical location of the monitored points on the bridge. The position of one given point is denoted (Support-id, Girder-id, Distance from support), where the parameters are defined as follows:

- **Support-id**: There are 52 monitored supports along the bridge. These are labeled S2-S30 on the southern part of the bridge, I-IX in the middle part and N0-N13 on the northern end.
- **Girder-id**: Each girder of the Göta River Bridge is labeled with a single character girder-id from A to G. The five girders in the middle, girder B, C, D, E and F are monitored.
- **Distance from support**: The distance in metres from the support. The resolution is 1cm.

Cracks have earlier been detected at support N7 and N9, girder A. Since girder A is not monitored, we choose to look at girder B. The location of the point analyzed in this thesis is (N7, B, 13.15).

3.3. Definition of Strain

Deformation in a material can be caused by many factors such as external applied loads, changes in temperature and irradiation effects. Strain is a non-dimensional measure of deformation, which represents the change in dimension per unit of original dimension. In Benham et al. (1996) *direct strain* for an unloaded uniform bar as shown in Figure 3.2 is defined as

$$\varepsilon = \frac{\delta}{L}$$

where L is the original length of a body and δ is the additional length after the tensile load F is applied to the body. If the deformation of the body leads to a compression, then δ is negative and represent the decrease in length. Strain can be divided into two parts, *thermal strain*, ε_T,

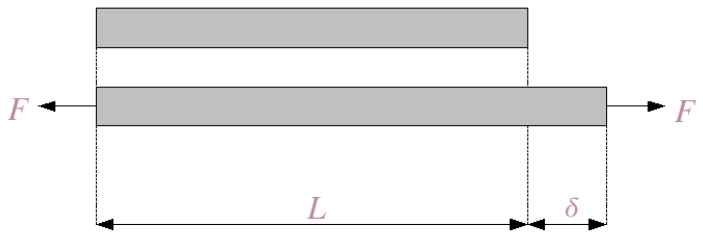


Figure 3.2: Deformation of a body due to tensile load.

and strain associated with stress, ε_σ. Thermal strain which is deformation due to temperature is defined as

$$\varepsilon_T = \alpha(T - T_0)$$

where $T - T_0$ is the change in temperature from a reference temperature T_0. α is a the *coefficient of linear thermal expansion*, a constant defining change in length per unit length for unit increase in temperature. Stress is the internal force per area unit and is denoted by σ. We have that

$$\varepsilon_\sigma = \frac{\sigma}{E}$$

where E is a constant related to the elasticity of the body. Thus

$$\varepsilon = \varepsilon_T + \varepsilon_\sigma = \alpha(T - T_0) + \frac{\sigma}{E} \tag{3.1}$$

The data from the Göta River Bridge are given in micro strain units. Given a reference length $L = 10cm$, an additional length $\delta = 0.1mm$ would correspond to a deformation of 1 micro strain. According to Glišić et al. (2007), the accuracy of the strain data is plus/minus 21 micro strain.

3.4. Strain, Temperature and Radiation Data

Strain data are available from 11.08.07 at 00.00Hrs to 02.10.07 at 22.00Hrs. There are 12 system readings in a day, one every second hour starting at 00.00Hrs. This makes up for a total of 636 observations. The first 48 observations (that is 4 days), are used as a burn-in period when calibrating forecast models. The observations in the period between 15.08.07 at 00.00Hrs to 01.10.07 at 22.00Hrs are divided into three overlapping datasets: Dataset 1 (D1) includes the first third of the data, the 16 days from 15.08.07 at 00.00Hrs to 31.08.07 at 22.00Hrs. Dataset 2 (D2) covers 32 days, from 15.08.07 at 00.00Hrs to 15.09.07 at 22.00Hrs. Finally, Dataset 3 (D3) includes all observations from 15.08.07 at 00.00Hrs to 01.10.07 at 22.00Hrs. The 12 remaining data readings from 01.10.07 at 00.00Hrs to 02.10.07 at 22.00Hrs are not used when calibrating the forecast models, but will be used later to test the quality of step ahead forecasts. The partitioning of the dataset is depicted in Figure 3.3. The motivation for operating with several datasets, is to monitor the predictive performance for different parts of the time series. It is not necessarily given that one model or one specific combination of discount factors is optimal throughout the full dataset. Examining different parts of the dataset will give us an overview of performance over time, and changes in the time series behavior. Performance for D3, which corresponds to the full dataset (excluding burn-in and the 12 last observations), is crucial when selecting a model, as the winning model for D3 is expected to be the best performing model.

Figure 3.4 (a) shows strain data from 11.08.07 at 00.00Hrs to 02.10.07 at 22.00Hrs. It looks like the data have a constant underlying trend, but as the series is quite volatile, a linear trend should not be dismissed without further investigation. The strain time series might have periodical

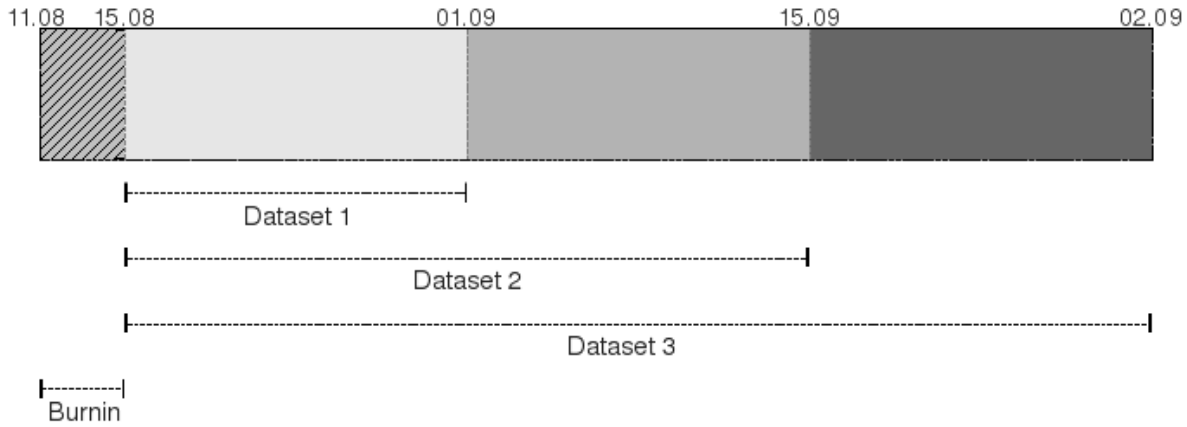
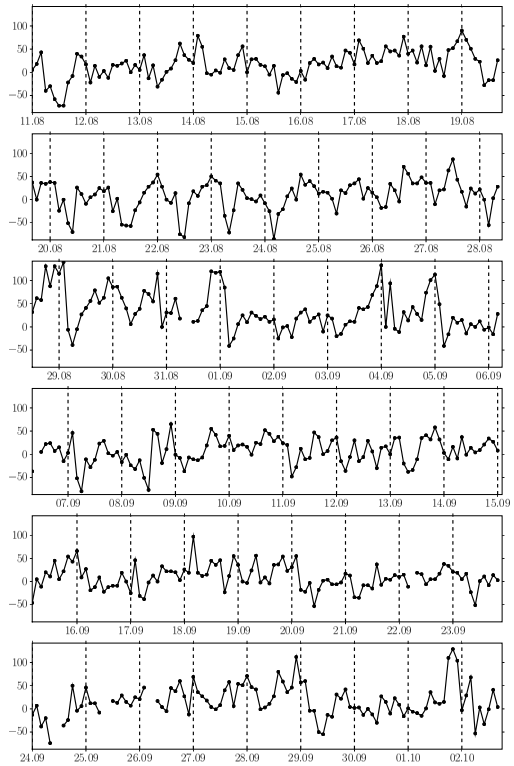


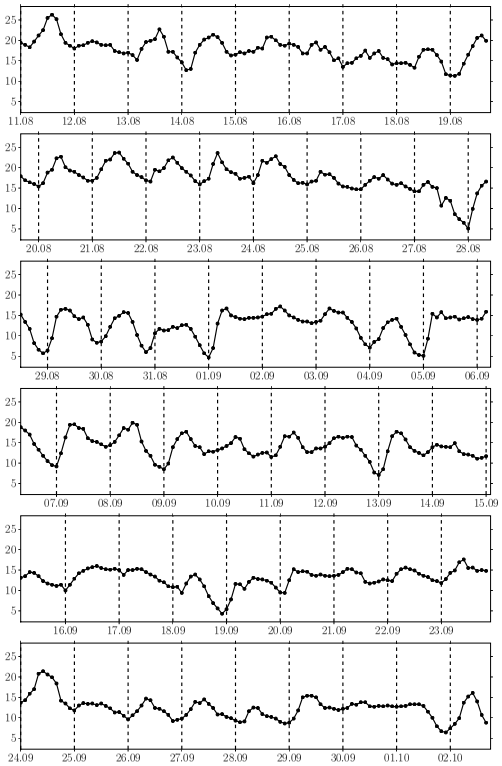
Figure 3.3: The dataset is divided into several parts to evaluate how models performs over time.

behavior. There is a seasonal tendency in the data, with a period of one day, and maximum values close to midnight. This implies that it might be appropriate to include a season component with period 12 when modeling the data. Figure 3.4 (b) shows temperature data for the same time period. As (3.1) implies, there seems to be a relation between the strain data and the temperature data. When the temperature decreases, the strain values increase. (3.1) actually claims the opposite - that strain values and temperature differences are positively correlated. The definition is, however, based on an unbraced body, while the bridge is a complex, supported body, which makes the local behavior more unpredictable. While some parts of the bridge expand as a result of high temperatures, other areas can be compressed. The apparent link between temperature and strain should be studied with a regression block model. Global radiation data from the period of interest are shown in Figure 3.4 (c). Global radiation measures direct solar radiation and diffuse sky radiation on the earth surface. As for the temperature data, it looks like the radiation data are negatively correlated to the strain data. We should however have in mind that radiation and air temperature are correlated. Nevertheless, radiation could be a potential explanatory variable and should be tested as well. NGI has provided the strain data, while SMHI (Swedish Metrological and Hydrological Institute) has contributed with both the temperature and the radiation data.

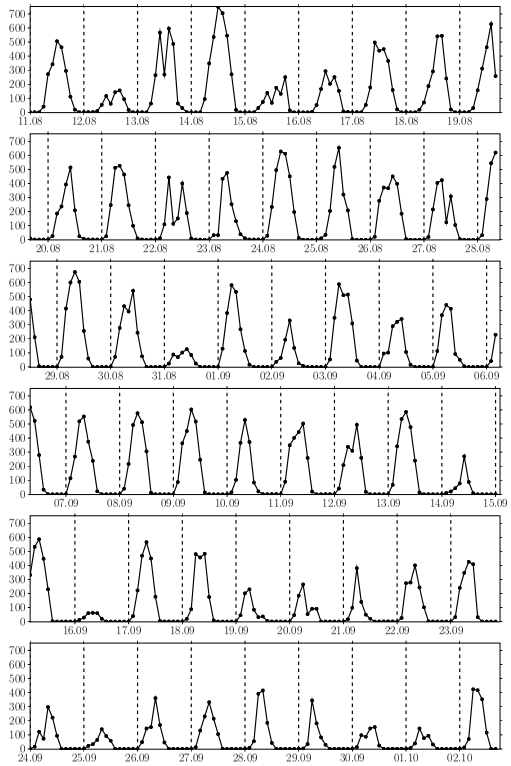
3.4.1. Missing Values. There are several missing values in the strain dataset. Handling missing values is straightforward in Bayesian time series analysis. As a part of the Bayesian time series dynamics, prior distributions are defined at each time step. If there is a missing value at a given point of time, the posterior distribution would be equal to the prior. There is no reason to adjust the prior distribution as no new information is available. Say for example that the observation Y_t is missing. Then West & Harrison (1997) defines $p(\boldsymbol{\theta}_t|D_t) = p(\boldsymbol{\theta}_t|D_{t-1})$, where $p(\boldsymbol{\theta}_t|D_t)$ and $p(\boldsymbol{\theta}_t|D_{t-1})$ is respectively the posterior and the prior distribution for time t. Normally the uncertainty of the posterior distribution is lower than the uncertainty of the prior, thus missing values lead to some added uncertainty in the model.



(a) Strain data from (N07,B,13.15), in micro strain.



(b) Temperature data from Gothenburg, in °C



(c) Global radiation data from Gothenburg, in Wh/m^2

Figure 3.4: Raw data

CHAPTER 4

Software

While West & Harrison (1997) gives a complete overview of the DLM framework, the main focus in Pole et al. (1994) is to apply the theory to univariate DLMs. BATS, a software package developed to analyze time series data in a Bayesian context, is included in Pole et al. (1994). BATS offers a variety of functions for constructing block structured DLMs, forecasting and analyzing time series data. According to Tvete (2000) and Myhre (1997) the software does however have many flaws and deficiencies, which makes it difficult to use. A critical task when building DLMs is to identify discount factors which give the optimal balance between flexibility and stability. BATS does not have a function for finding the optimal combination of discount factors, it has to be done by testing out one combination at a time. With up to three discount factors ranging between 0 and 1, this process is of course time-consuming, and as a result it limits the span of tested combinations. This is the main reason why we decided to make our own software. It also opens up the possibility to implement other functions that are not included in BATS. The software is written in Python, a high-level scripting language well-suited for scientific applications. The full source code is available at: http://folk.uio.no/idakso/python/.

The Python program has more or less the same functions as BATS, but with a few variations. First of all BATS uses a discount strategy for the observation variance, similar to the approach used for state parameter variances described in Section 2.9.4. Using this strategy allows for a non-constant observation variance. For the bridge data it is reasonable to assume a constant observation variance, and thus the approach introduced in Section 2.5 is used in the Python program. Second, for state parameter variances BATS uses the component discount method defined in (2.6), while we chose to use the strategy defined in (2.7) to ensure that the correlations of the state parameters are not changed due to discounting. These two distinctions may lead to some small differences in both state and observation variances, and as a consequence also to discrepancies in estimated parameter and forecast values. Analyses from the two softwares should however not differ much, as will be illustrated in the following example.

EXAMPLE 4.1. *To compare results from BATS and our Python program, and to make sure that the latter works properly, both programs will be used to analyze a synthetic dataset constructed for this purpose. The dataset is put together as a sum of four components: a sine curve with period 12, a constant zero trend, a regression term and noise. The data is constructed in this manner to easily be able to recognize the data structure with a component DLM. More specifically*

Y_t is constructed as follows:

$$Y_t = 0 + 4\sin\left(\frac{\pi}{6}t\right) + 5X_t + \epsilon_t,$$

where $t = 1, 2, ..., 47, 48$. The terms represent trend, season, regression and noise respectively. $\mathbf{X} = (X_1, X_2, ..., X_{48})'$ is a random set of regression variables drawn from a uniform distribution and ϵ_t is drawn from a standard normal distribution.

To analyze the synthetic data, the DLM used is based on the known data structure:

$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V_t],$$

where Y_{T_t} is a constant trend block, Y_{S_t} is a single harmonic with period 12 and Y_{R_t} is a simple linear regression on X_t. Constant observation variances are assumed using the Python program, thus $V_t = V$. The block components are defined by

$$\begin{array}{lll} \boldsymbol{\theta}_{T_t} = \mu_t, & \mathbf{F}_{T_t} = \mathbf{F}_T = 1, & \mathbf{G}_{T_t} = \mathbf{G}_t = 1, \\ \boldsymbol{\theta}_{S_t} = (a_t, b_t)', & \mathbf{F}_{S_t} = \mathbf{F}_S = (1,0)', & \mathbf{G}_{S_t} = \mathbf{G}_S = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}, \\ \boldsymbol{\theta}_{R_t} = \beta_t, & \mathbf{F}_{R_t} = X_t, & \mathbf{G}_{R_t} = \mathbf{G}_R = 1 \end{array}$$

where $\omega = \frac{2\pi}{p}$ with $p = 12$. As the underlying dynamics are known, we construct the initial prior:

$$\mathbf{m}_0 = \begin{pmatrix} 0 & 0 & 4 & 5 \end{pmatrix}', \qquad \mathbf{C}_0 = \mathbf{I}_4, \qquad n_0 = 1, \qquad S_0 = 1$$

The first 0 term in $\mathbf{m}_0$ is the prior expected level μ_1 and the last term 5 is the prior expected value for the regression coefficient β_1. The sequence in the middle, $(0, 4)$, is the prior expected values for the seasonal parameters (a_1, b_1). Since we know the underlying seasonal effects ϕ, appropriate priors for (a_1, b_1) is easily found using (2.4). To reflect low uncertainty around the state parameters, all state parameter variances are set to 1. Since BATS allows for non-constant observation variance, an initial prior is appointed for V_0 rather than S_0. V_0 is set to 1 in BATS, as this corresponds to the variance of the error term ϵ_t. Similarly, S_0 was set to 1, as S_0 represents the prior point estimate for V. The discounts factors for the component blocks are set to $\delta_T = 0.90$, $\delta_S = 0.95$ $\delta_R = 0.98$. Figure 4.1 shows 1-step predictions, estimated level, seasonal effects, regression coefficient and prediction standard deviations for BATS and for our alternative software. We see from Figure 4.1 that the two programs have identical results for the level, seasonal effects and regression coefficient. However the estimated prediction standard deviation levels are slightly different. This is not very surprising, since the prediction variance is defined by $Q_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + V_t$, and the programs have different approaches for estimating V_t and $\mathbf{R}_t$ (the latter due to discounting). It does not seem to be of great importance for the 1-step predictions. The prediction interval is slightly broader for small t-values using the Python program. Otherwise there are no evident variations. It is interesting to observe that the prediction standard deviation seems more correct using the Python program with the constant variance assumption. The correct

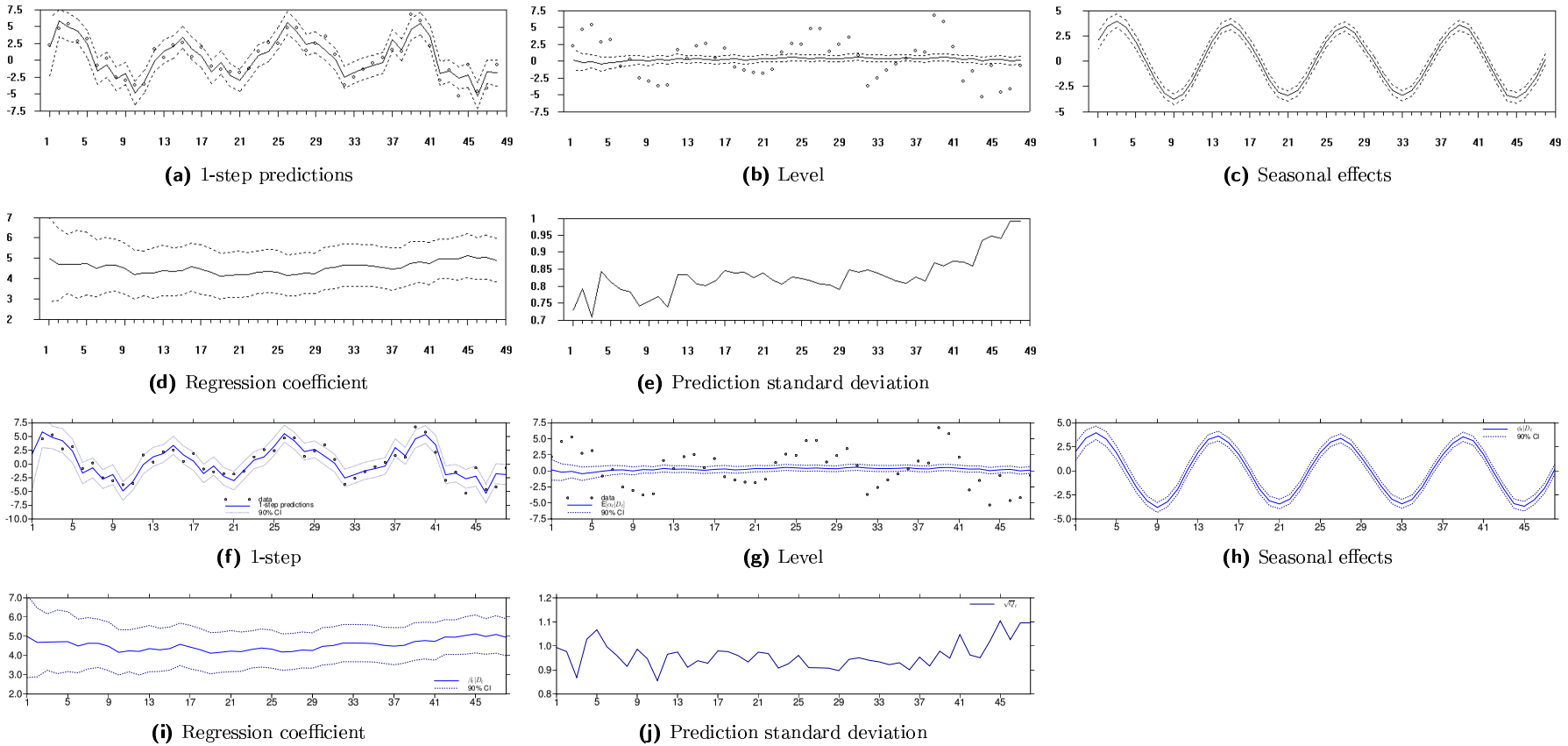


Figure 4.1: *(a)-(e): BATS results (f)-(j): Results from Python program*

underlying observation variance V or V_t is 1. Since the prediction standard deviation includes both observation and parameter uncertainty, it should be at least 1. Even though the standard deviations differ to some degree, the dynamics of the the two are similar.□

CHAPTER 5

Model Building

In this chapter we will apply the DLM theory introduced in Chapter 2 to the strain time series data from the Göta River Bridge. Different models are explored and discussed, and challenges related to regression models and step ahead forecasting are addressed. Numerical summaries of predictive performance are used as a basis for comparing models and possibly single out a winning model. Three basic measures are presented in the following section.

5.1. Model Criteria

West & Harrison (1997) presents three commonly used measures of forecast accuracy: mean absolute deviation (MAD), root mean square error (RMSE) and log likelihood (LLH). The two first criteria measure predictive accuracy based on point forecast errors. The *mean absolute deviation* is defined as

$$\text{MAD} = \sum_{t=1}^{T} |e_t|/T,$$

where e_t, referred to as errors or residuals, represents the difference between the 1-step predicted value $\text{E}(Y_t|D_{t-1})$, and the real observed value Y_t. The *root mean square error* is given by

$$\text{RMSE} = \sqrt{\sum_{t=1}^{T} e_t^2/T}$$

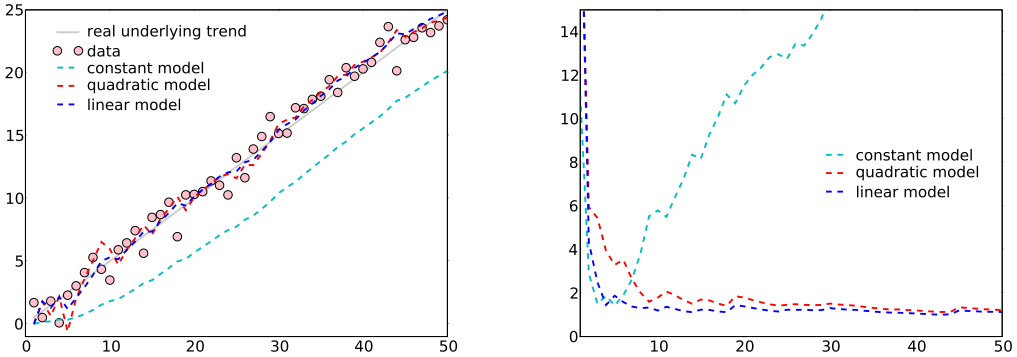
The *log likelihood* measures the quality of the observed predictive density $p(Y_t|D_{t-1})$:

$$\text{LLH} = \sum_{t=1}^{T} \log(p(Y_t|D_{t-1})),$$

where Y_t is the observed data, and $p(Y_t|D_{t-1})$ is the probability of the real observed Y_t according to the predictive distribution (and not the predictive distribution itself, even though the terminology is the same). By maximizing the LLH, predictive distributions with accurate expected values combined with low variances are favored. When comparing LLH for datasets of different sizes, it is convenient to use the average log likelihood contribution per time unit, LLH/T.

EXAMPLE 5.1. *In the following example the synthetic data set has a linear trend and standard normally distributed noise. Three different models, a 1st, 2nd and 3rd order polynomial trend, are evaluated to compare model performance. A* 0.95 *discount factor is used for all models. Figure 5.1 (a) shows 1-step predictions for the three models. The constant trend (1st order) predictions*

are consistently below the data, as its only state parameter is the level. Both the linear and the quadratic model are doing well, but the quadratic is somewhat less stable, since it has a redundant third parameter modeling noise. From Figure 5.1 (b) we see that the prediction variance for the 1st order model increases over time, as the predictions never hit target. Both the quadratic and linear model prediction variances converge after some time, the latter converging faster.



(a) 1-step predictions **(b)** Prediction variance

Figure 5.1: *(a) 1-step predictions and (b) prediction variances for 1st, 2nd and 3rd order polynomial trend models applied to data with an underlying linear trend.*

To avoid evaluating a volatile burn in period, the first 10 predictions are not taken into account. The following results are achieved.

Model	*RMSE*	*MAD*	*LLH*
1st order	*4.76*	*4.64*	*-145.59*
2nd order	*1.07*	*0.87*	*-61.54*
3rd order	*1.17*	*0.90*	*-65.87*

Table 5.1: *Results for 1st, 2nd and 3rd order trend models.*

All criteria confirm that the 2nd order trend has the best predictive performance of the three models. It minimizes both the RMSE and the MAD, and maximizes the LLH. The 3rd order polynomial trend has somewhat poorer results due to overfitting.

The performance measures are also used to compare alternative discount factors. In our applications, we have decided to find the optimal discount factor 'brute force' - by iterating through a certain range of discount factors. For each combination of discount factors (one for each block), RMSE, MAD and LLH are calculated. It is often not possible to find one combination which optimizes all criteria. The combination that gives the best RMSE score might not necessarily give optimal MAD and LLH scores. The different criteria have three different functions, which are all important for a good model and good predictions, thus we decided to weigh all criteria

equally when deciding on optimal discounts factors. Each combination of discount factors is given a score based on how well it does according to the three criteria. If a given combination of discount factors gives the best RMSE of all combinations, the second best MAD and the third best LLH, then the overall score is $1 + 2 + 3 = 6$. The combination that minimizes the overall score is chosen as the optimal discount factors for the evaluated model. When choosing between competing models that are hard to separate in terms of performance, the models' RMSE scores are emphasized. Good short term predictions are important in the bridge case, and small deviations are not as critical as large ones. RMSE is a good measure for prediction performance, and it punishes large errors.

All suggested models are evaluated for all three datasets described in Chapter 2. The purpose of doing this is to achieve an understanding of how the model performs throughout the different parts of the dataset. The model performance for the full dataset, D3, is decisive when ultimately choosing an optimal model.

5.2. Model Suggestions

5.2.1. Model 1: Trend and Season. The first model suggestion for the strain data is a simple trend and season model:

$$Y_t = Y_{T_t} + Y_{S_t} + \nu_t, \qquad \nu_t \sim N[0, V]$$

When the data were explored in Section 3.4, we found indications of a seasonal pattern with period 12 (one day) in the strain data. Such behavior can be modeled with a single harmonic:

$$\begin{aligned} \boldsymbol{\theta}_{S_t} &= (a_t, b_t)' \\ \mathbf{F}_{S_t} &= \mathbf{F}_S = (1, 0)' \\ \mathbf{G}_{S_t} &= \mathbf{G}_S = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}, \end{aligned} \tag{5.1}$$

where $\omega = \frac{2\pi}{p}$ with $p = 12$.

It is not obvious whether a constant or a linear trend gives the best fit for the data, so both alternatives are explored. For Model 1a, a constant trend is used:

$$\begin{aligned} \boldsymbol{\theta}_{T_t} &= \alpha_t \\ \mathbf{F}_{T_t} &= \mathbf{F}_T = 1 \\ \mathbf{G}_{T_t} &= \mathbf{G}_T = 1 \end{aligned} \tag{5.2}$$

while the trend is linear for Model 1b:

$$\boldsymbol{\theta}_{T_t} = (\alpha_t, \mu_t)'$$
$$\mathbf{F}_{T_t} = \mathbf{F}_T = (1, 0)'$$
$$\mathbf{G}_{T_t} = \mathbf{G}_T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

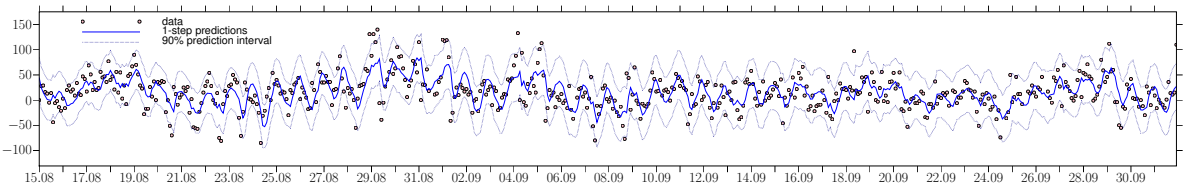
Both models are tested on all three datasets described in Section 3.4. The initial variance for all parameters are chosen adequately large, so that the initial prior assumptions are non-informative in practice. The first 48 observations (corresponding to 4 days) are used as a burn-in period and are not taken into account when evaluating models and choosing optimal discount factors. (Neither increasing the burn-in period, nor the initial variance has significant effects on the results.)

Dataset	Model	Burn-in	RMSE	MAD	LLH/T	δ_T	δ_S
D1	M1a	48	28.39	22.01	-7.34	0.78	0.88
D1	M1b	48	28.95	22.41	-7.35	0.91	0.87
D2	M1a	48	29.56	22.87	-7.45	0.80	0.94
D2	M1b	48	30.18	23.47	-7.48	0.91	0.93
D3	M1a	48	28.24	21.65	-7.43	0.77	0.98
D3	M1b	48	29.0	22.36	-7.45	0.90	0.95

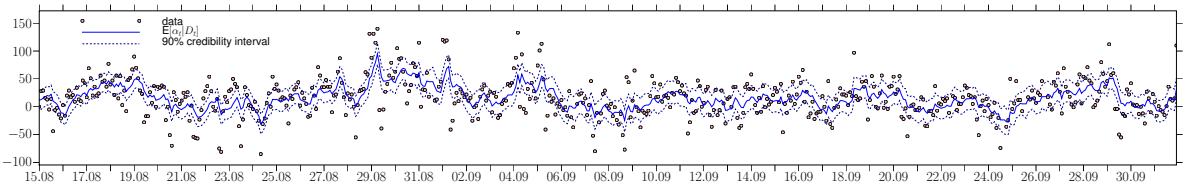
Table 5.2: Results for Model 1a and Model 1b.

According to the results given in Table 5.2, Model 1a with a constant trend block has the best performance for all three datasets. Hence Model 1b with the linear trend block is discarded. For Model 1a, the discount factor for the season block, δ_S increases with the size of the dataset, which indicates that the underlying seasonal effects are more stable towards the end of the full dataset. The discount factor for the trend block, δ_T, does not change very much, varying from a minimum 0.77 for D3 to a maximum 0.80 for D2. These are quite low values for a trend discount factor, and it is a sign that the model is inadequate. Both RMSE and MAD show that Model 1a performs better for D3 than D1 and D2. The best LLH score is obtained for D1. The predictive performance for D2 is poorer according to the LLH scores, while there is an improvement in the score for D3. Have in mind that D3 includes both D1 and D2, so the numbers are not directly comparable, but we can look for positive and negative changes in the criteria from one dataset to the next. All in all the model seems to perform better for the first and the last part of the full dataset, and somewhat poorer for the middle part.

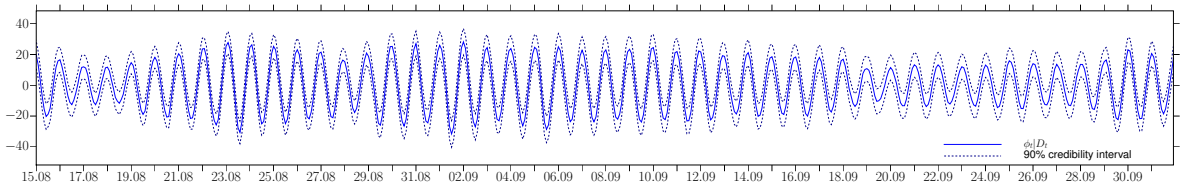
In Figure 5.2, results from Model 1a applied to D3 are plotted. Figure 5.2 (a) shows the data plotted with 1-step predictions and a corresponding 90% prediction interval. The fit is not too bad, but quite a few observations do not lie within the prediction interval. Studying the air temperatures in Figure 3.4 (b), it is apparent that most of these observations have been done



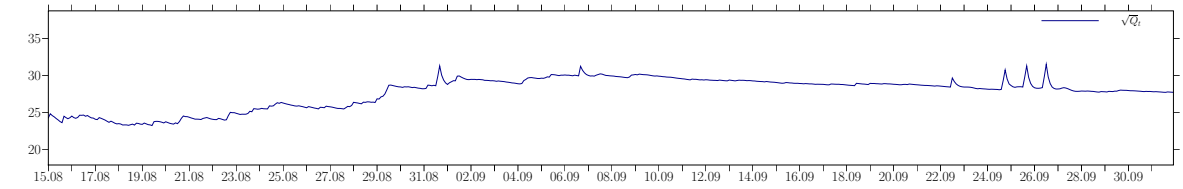
(a) 1-step predictions



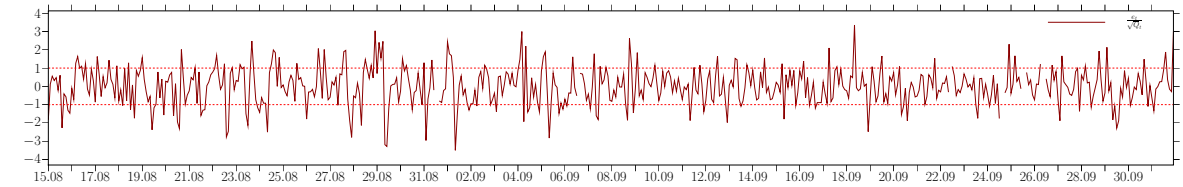
(b) Level



(c) Seasonal effects



(d) Prediction standard deviation



(e) Standardized errors

Figure 5.2: M1a, D3, $\delta_T = 0.77$, $\delta_S = 0.98$

when there are large fluctuations in temperature. This suggests that temperature should be used as an explanatory variable.

In Figure 5.2 (b), we observe the expected posterior level parameter, $\mathrm{E}[\alpha_t|D_t]$, with a 90% credibility interval. The level is very volatile, and to a certain degree it compensates for more extreme strain observations. This leads to increased uncertainty on the level parameter, which in turn also contributes to greater prediction uncertainty, as we can see in Figure 5.2 (d). The volatility in the trend parameter is a sign of an inadequate model, which again suggests that we should introduce a regression block. The small peaks in the prediction standard deviation are caused by missing values.

Figure 5.2 (c) shows the posterior expected seasonal effects, $\mathrm{E}[\phi_t|D_t]$, with a corresponding 90 % credibility interval. There are variations over time, but these are slow, as the 0.98 discount factor for the season block does not allow the parameter to adjust very fast. Comparing to temperatures in Figure 3.4 (b), the seasonal effects seem related to air temperatures. When the temperatures are stable, as from 14.09 to 18.09, the amplitude of the seasonal effects decreases slowly. In more turbulent periods, as from 18.08 to 24.08, the amplitude increases. Again this suggests that temperatures affect strain values.

The standardized errors, $e_t/\sqrt{Q_t}$, are plotted in Figure 5.2 (e). It is assumed that $e_t|D_{t-1} \sim T_{n_{t-1}}[0, Q_t]$, hence the standardized errors, $e_t/\sqrt{Q_t}|D_{t-1}$, should have a $T_{n_{t-1}}[0,1]$ distribution. The assumption seems reasonable studying Figure 5.2 (e) (a more thorough analysis of this assumption will be done for our final model). A concern however is that there are several subsequent, correlated errors especially noticeable in the period from 27.08 to 29.08.

5.2.2. Model 2: Temperature as a regressor. In the second model suggestion, we introduce a regression block, with temperature as an explanatory variable. Even though the additional block probably explains a lot of the seasonal pattern seen in Model 1a, there might still be remaining seasonal behavior in the data. Hence it is necessary to try out two different model versions - with and without a season block. Model 2a includes both a trend, season and regression block:

$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} is a constant trend as in (5.2) and Y_{S_t} is a single harmonic with period 12, as defined in (5.1). The regression block is given by

$$\begin{aligned} \boldsymbol{\theta}_{R_t} &= \beta_t \\ \mathbf{F}_{R_t} &= TMP_t \\ \mathbf{G}_{R_t} &= \mathbf{G}_R = 1, \end{aligned}$$

where TMP_t is the air temperature at time t. Model 2b only has a trend and a regression block:

$$Y_t = Y_{T_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} and Y_{R_t} are defined as for Model 2a.

The results implementing Model 2a (M2a) and Model 2b (M2b) are given in Table 5.3. The burn-in period is 48 and the initial variances are chosen as previously to make the initial information assumptions non-informative. These are default values for all models. Figure 5.3 displays the performance scores for the new models and Model 1a. It is evident that both Model 2a and Model 2b performs considerably better than the first model suggestion. Model 2a with the season block does better than Model 2b for all datasets according to all criteria. Model 2a does better for D1 and D3 than for D2, as was the case for Model 1a. This implies that Model 2a

Dataset	Model	Burn-in	RMSE	MAD	LLH/T	δ_T	δ_S	δ_R
D1	M2a	48	24.22	17.93	-7.02	0.88	0.97	0.89
D1	M2b	48	25.33	19.21	-7.15	0.86	-	0.88
D2	M2a	48	24.91	18.54	-7.10	0.90	0.99	0.91
D2	M2b	48	25.90	19.53	-7.21	0.89	-	0.91
D3	M2a	48	24.06	18.12	-7.07	0.91	0.99	0.92
D3	M2b	48	24.73	18.71	-7.17	0.92	-	0.93

Table 5.3: Results for Model 2a, Model 2b

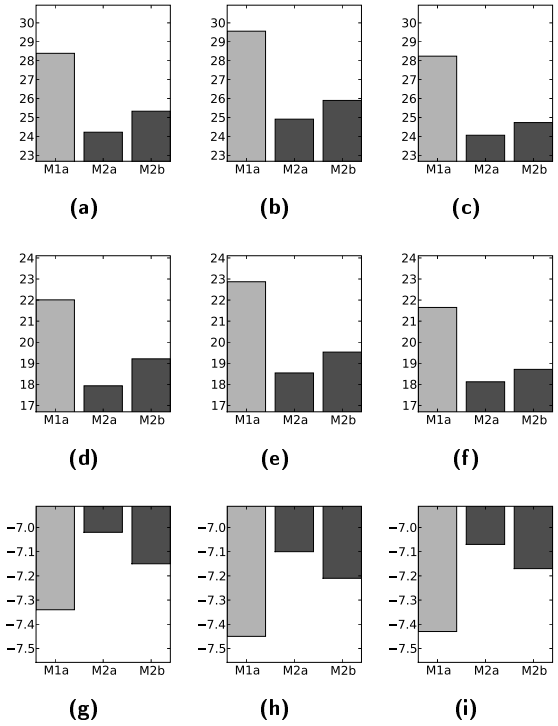
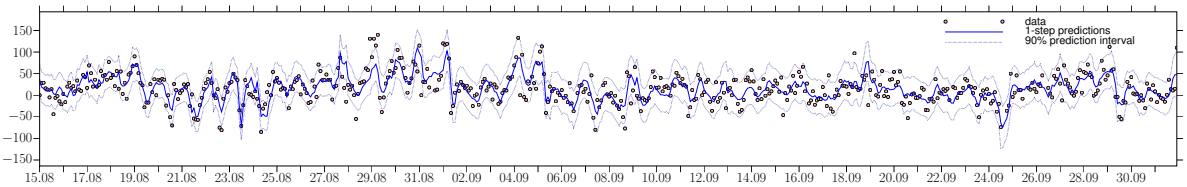


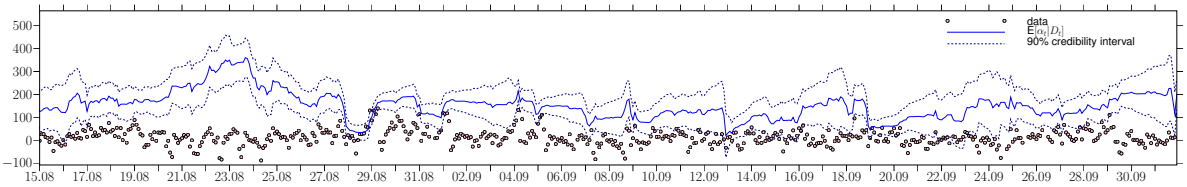
Figure 5.3: Performance scores for models M1a, M2a and M2b. (a)-(c): RMSE scores for D1, D2 and D3. (d)-(f): MAD scores for D1, D2 and D3. (g)-(i): LLH scores for D1, D2 and D3.

also performs better for the first and last part of the full dataset. From Table 5.3, we can see that the discount factors for all blocks are quite high, especially the season discount factor, δ_S. The discount factors generally increase as the dataset grows, but are quite stable.

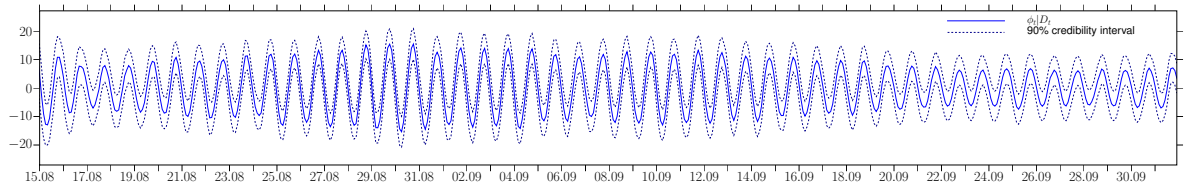
Figure 5.4 shows the results from applying Model 2a to D3. Studying the 1-step predictions in Figure 5.4(a) it is easy to see that Model 2a performs better than Model 1a. Fewer observations are outside the prediction interval than for Model 1a. But nevertheless there are some deviations, such as the series of quite extreme observations from 27.08 to 29.08. Model 1a actually has a better fit for these values. This can be explained by the low trend discount factor used for Model 1a, which allowed the level parameter to follow the data. From Figure 3.4(b) we see that there are great variations in temperature in the same period. This is often the case where the 1-step



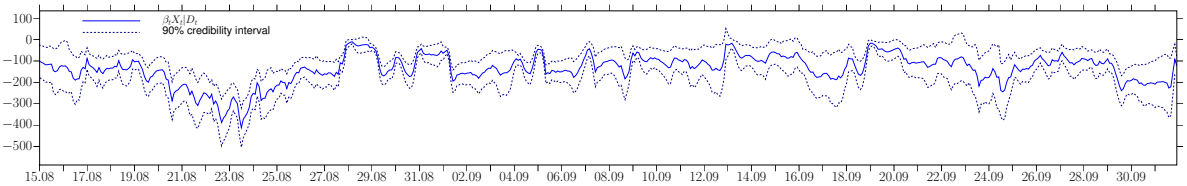
(a) 1-step predictions



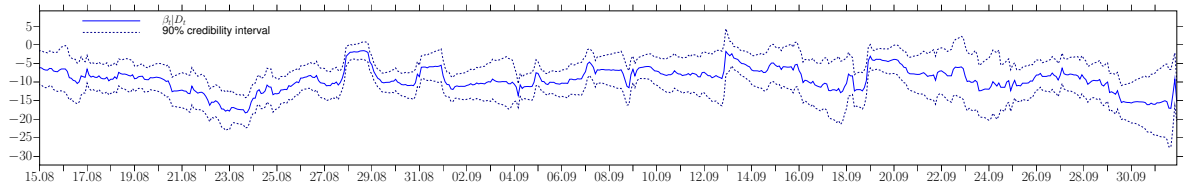
(b) Level



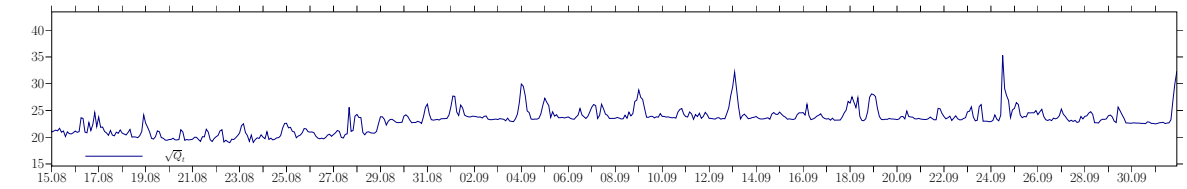
(c) Seasonal effects



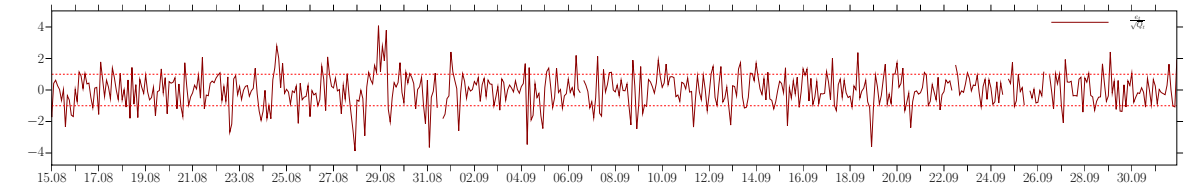
(d) Effect of temperature



(e) Regression coefficient temperature



(f) Prediction standard deviation



(g) Standardized errors

Figure 5.4: M2a, D3, $\delta_T = 0.91$, $\delta_S = 0.99$, $\delta_T = 0.92$

prediction does not do very well. It leads us to believe that temperature variations might be relevant as an explanatory factor, in addition to temperature itself.

The level parameter in Figure 5.4(b) lies high above the underlying level in the data. The level is very volatile with several sudden jumps, which in time induces a lot of uncertainty on the parameter, and also to jumps in the prediction standard deviation in Figure 5.4(f). (Some jumps are explained by missing values). The explanation for the off-level is found in Figure 4(d), which displays the effect of temperature. Comparing the temperature effect to the level parameter, it is clear that the level is confounded with the behavior of the temperature effect. The temperature is given in an arbitrary scale, thus the average contribution is positive. To achieve a level independent of the regression effects, the average contribution from the regression block must be zero, hence the temperature must be shifted to have a zero mean.

Comparing the seasonal effects for Model 2a in Figure 5.4(c) and for Model 1a, we can see that the amplitude of the latter is in general smaller (notice that the scales of the y-axes are not the same for the two figures). Since temperature explains a lot of the seasonal behavior seen earlier, this is expected.

Even though there are some jumps in the prediction standard deviation plotted in Figure 5.4(f), the level in generally lower than for Model 1a. Figure 5.4(g) shows the standardized errors. The errors are visibly smaller than for Model 1a, but there are still correlations in the errors.

5.2.3. Models 3 and 4: Shifting Temperature. Analyzing Model 2a, we discovered that it is necessary to shift the temperature variable, and this will be tested in Models 3 and 4. It is evident from the previous section that Model 2a, which included a season block, performed better than Model 2b. There is no reason to believe that shifting the regression variables will make the season block superfluous, so Models 3 and 4 are given as:

$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} is a trend block with a constant level as defined in (5.2), and Y_{S_t} is a season block with a single harmonic of period 12, as in (5.1).

In addition to shifting the parameters, it could be useful to standardize them as well, for better interpretation of the regression coefficients. For a static analysis that would be straightforward, but when working in a dynamical context it will complicate matters further, so we will not scale the temperatures. In Pole et al. (1994) shifting the regression coefficients is simply solved by doing a static shifting for the given regression variables. The same approach is used in Model 3, with the regression block:

$$\begin{aligned} \boldsymbol{\theta}_{R_t} &= \beta_t \\ \mathbf{F}_{R_t} &= TMP_t^{\#} \\ \mathbf{G}_{R_t} &= \mathbf{G}_R = 1, \end{aligned}$$

where $TMP_t^{\#} = TMP_t - \sum_t TMP_t/T$, and T is the number of observations in the dataset. When analyzing a static dataset, this solution is reasonable. The problem appears when the dataset is dynamic. This will be the case for the bridge data, when doing real-time analysis. An alternative approach is to perform a dynamic shifting, where temperature data at each point of time are shifted individually. Instead of using a static mean value, a moving average is subtracted from the given temperature. This also works as a filter, as it removes slow variations in temperature over time. Using a small time window will give an effective filtering, while more details will be preserved using a larger window. Trying out different alternatives, we have decided to use a symmetric window around the the given datapoint. The regression block for Model 4 is defined as

$$\begin{aligned}\boldsymbol{\theta}_{R_t} &= \beta_t \\ \mathbf{F}_{R_t} &= TMP_t^* \\ \mathbf{G}_{R_t} &= \mathbf{G}_R = 1,\end{aligned}$$

where $TMP_t^* = TMP_t - \frac{1}{2k+1}\sum_{i=t-k}^{t+k} TMP_i$. Both the dynamic and the static shifting methods require information about future temperatures. It is assumed that this can be obtained, preferably as a statistical model with uncertainties that can be included in the model. Experimenting with different window sizes, $k = 186$ gave the best results. This corresponds to plus/minus 15.5 days. The results from the two models are given in Table 5.4.

Dataset	Model	Burn-in	RMSE	MAD	LLH/T	δ_T	δ_S	δ_R
D1	M3	48	24.49	18.22	-6.96	0.90	0.99	0.89
D1	M4	48	24.74	18.61	-6.95	0.92	0.99	0.90
D2	M3	48	24.96	18.62	-7.06	0.92	1.0	0.91
D2	M4	48	24.94	18.85	-7.05	0.93	1.0	0.90
D3	M3	48	24.10	18.21	-7.04	0.93	1.0	0.92
D3	M4	48	24.13	18.36	-7.04	0.93	1.0	0.91

Table 5.4: Results for Model 3 and Model 4.

Figure 5.5 shows the predictive performances for Models 3 and 4, as well as Model 2a. There are only marginal differences in the results for Models 3 and 4. Which model performs best varies from dataset to dataset, and what criterion is used. According to the RMSE and MAD scores, Model 3 has the overall best performance of the two. The exception is for D2 where Model 4 has a better RMSE score. Model 4 has the best LLH scores for D1 and D2, while Models 3 and 4 perform equally for D3. Comparing the results to Model 2a, RMSE and MAD are slightly better for Model 2a while LLH is maximized for Models 3 and 4.

All discount factors increase as the size of the dataset grows, which implies less information loss later in the dataset. The season block has a discount factor $\delta_S = 1.0$, thus the seasonal effects are very stable. The discount factors for the trend and regression block are also quite high.

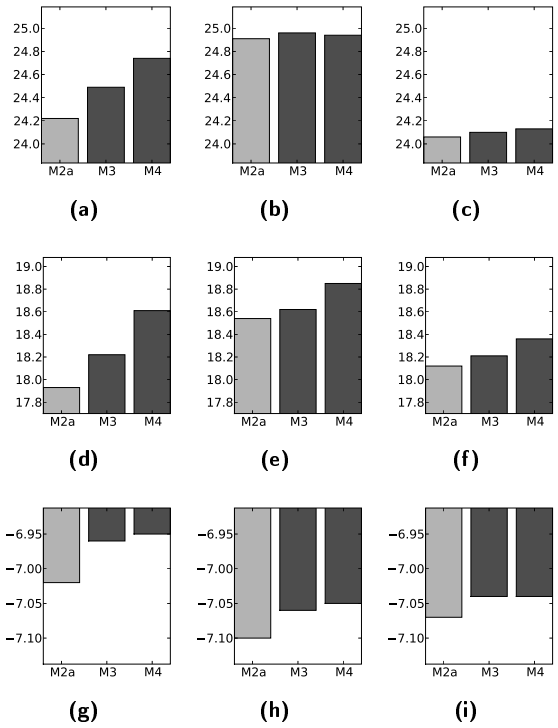
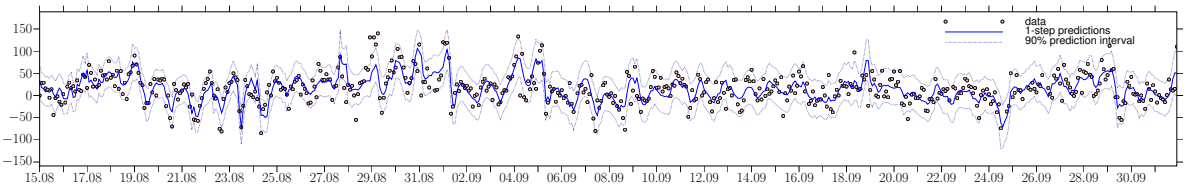


Figure 5.5: Performance scores for models M2a, M3 and M4. (a)-(c): RMSE scores for D1, D2 and D3. (d)-(f): MAD scores for D1, D2 and D3. (g)-(i): LLH scores for D1, D2 and D3.

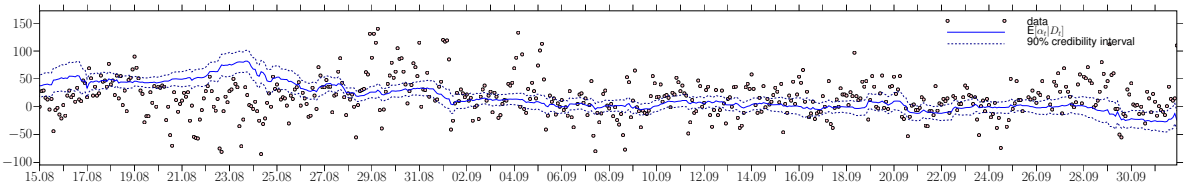
Figure 5.6 shows results applying Model 3 to D3. The 1-step predictions for Model 3 in Figure 5.6 (a) and for Model 2a are quite similar. Comparing the level parameter for Model 3 in Figure 5.6 (b) and to the level for Model 2a, it is clear that shifting the temperature had the desired effect. The level parameter is much more representative for the underlying level of the time series. The uncertainty of the level has also drastically decreased. But the level parameter still has a decreasing trend, which is not apparent in the raw strain data. The temperature data on the other hand decreases from August to October, thus the mean static shifted temperature for the first half is positive, while it is negative for the last half. The regression coefficient β_t in Figure 5.6 (e) is negative, thus positive temperatures give negative effects. The level parameter compensates for the increasing mean effect of the regression component, and thus it is indirectly affected by the slow decrease in the temperature.

Compared to Model 2a, the regression coefficient shown in Figure 5.6 (e), is less volatile and has less uncertainty. A stable regression coefficient with low variance is optimal. Short term forecasts can be good even though the uncertainty is high, while medium- and long term predictions are often poorer.

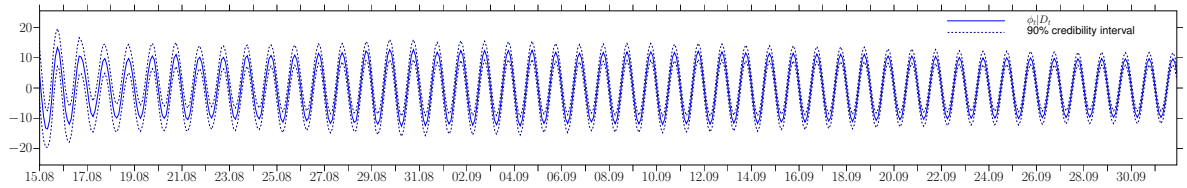
The seasonal effects for Model 3 in Figure 5.6 (c) are more stable than for Model 2a. The discount factors for both models are high - 0.99 for Model 2a and 1.0 for Model 3. Thus shifting the regression variables also had a stabilizing effect on the seasonal behavior.



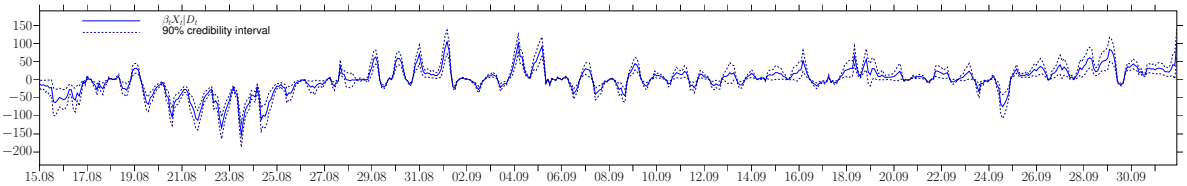
(a) 1-step predictions



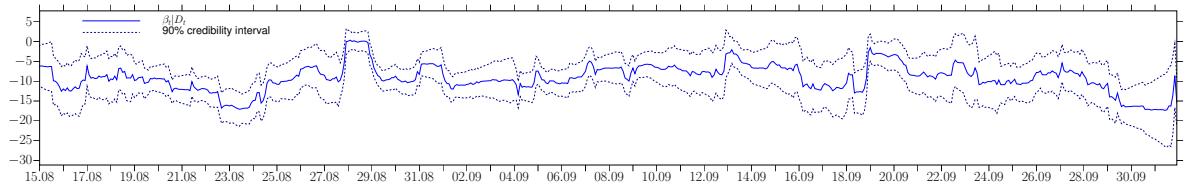
(b) Level



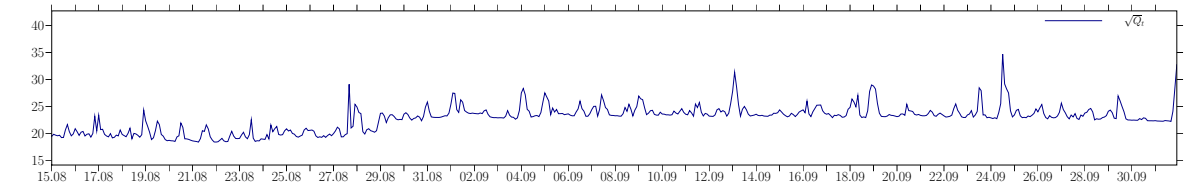
(c) Seasonal effects



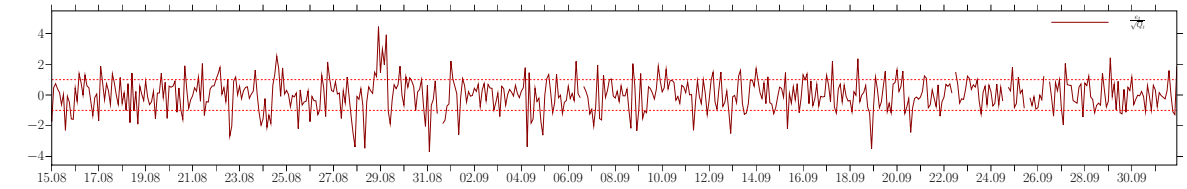
(d) Effect of temperature



(e) Regression coefficient temperature



(f) Prediction standard deviation



(g) Standardized errors

Figure 5.6: M3, D3, $\delta_T = 0.93$, $\delta_S = 1.0$, $\delta_T = 0.92$

Figure 5.6 (f) shows the prediction standard deviation for Model 3. There are still jumps similar to the ones in Model 2a, but the level of the standard deviation is slightly lower for Model 3.

The 1-step predictions for Model 4, shown in Figure 5.7 (a), have much resemblance to both Model 2a and Model 3. Comparing the level parameter for Model 4, shown in Figure 5.7 (b), to the corresponding for Model 3, we see that Model 4 has a more stable level with less uncertainty. There is no longer evidence that changes in the temperature affects the level parameter. Even though Model 4 is not the best model according to RMSE and MAD, the level parameter using dynamic shifting gives a more intuitive understanding of the underlying level of the data.

The seasonal effects for Model 4 shown in Figure 5.7 (c) are very similar to the effects for Model 3. Thus the two shifting methods do not seem to affect the seasonal parameters differently. The regression coefficient β_1 in Figure 5.7 (e) also seems more or less unchanged from Model 3. The effect of the temperature in Figure 5.7 (d), is however more stable than for Model 3, as for the level parameter.

Figure 5.7 (f) shows the prediction standard deviation for Model 4, which is on the same level as for Model 3. The standardized errors for Model 4 shown in Figure 5.7 (g) also have the same structure as for Model 3, and the main concern is still the subsequent, positively correlated errors. The model still predicts poorer where the data are more unstable, which might be caused by quick changes in temperature. To address this, it could be interesting to include temperature difference as an additional explanatory variable.

5.2.4. Models 5, 6 and 7: Regression with Temperature Difference. The results from earlier models suggest that temperature difference should be introduced as an explanatory variable as well as the temperature itself. Hopefully this move will solve the problem with correlated errors. The temperature difference at time t is defined as $TD_t = TMP_t - TMP_{t-1}$. Models 5a, 6a and 7a are defined as follows:

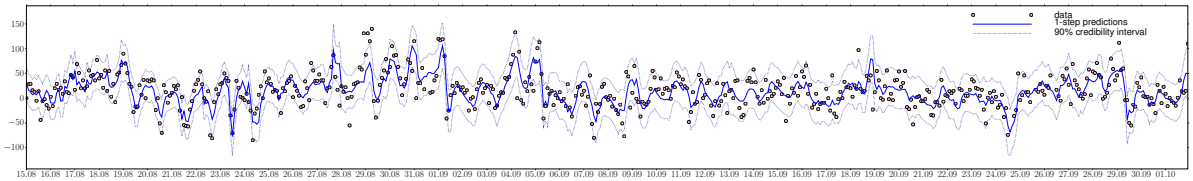
$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} is a trend block with a constant level as defined in (5.2), and Y_{S_t} is a season block with a single harmonic of period 12 as defined in (5.1). Models 5b, 6b and 7b are identical to respectively Models 5a, 6a and 7a, only without the season block:

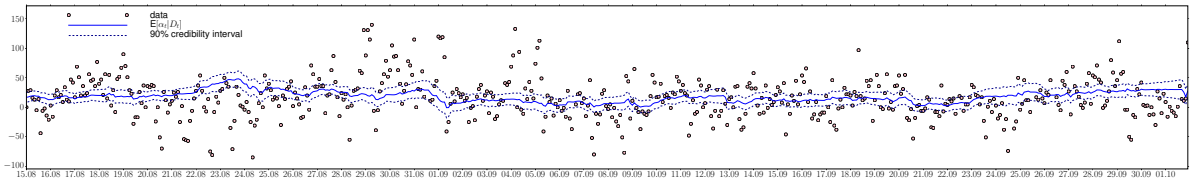
$$Y_t = Y_{T_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V]$$

For Models 5a and 5b the regression block is given by

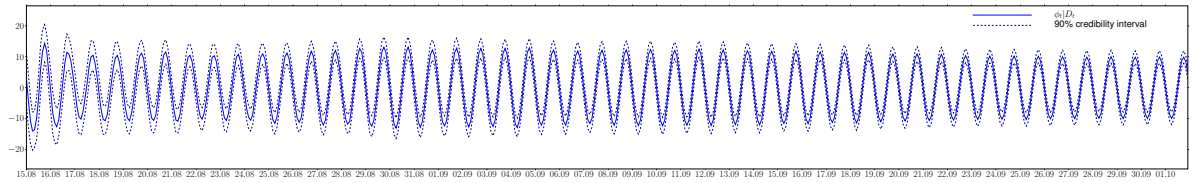
$$\boldsymbol{\theta}_{R_t} = (\beta_{1,t}, \beta_{2,t})'$$
$$\mathbf{F}_{R_t} = (TMP_t, TD_t)'$$
$$\mathbf{G}_{R_t} = \mathbf{G}_R = \boldsymbol{I}_2$$



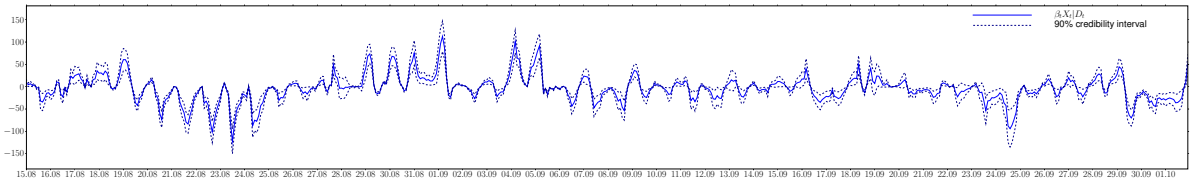
(a) 1-step predictions



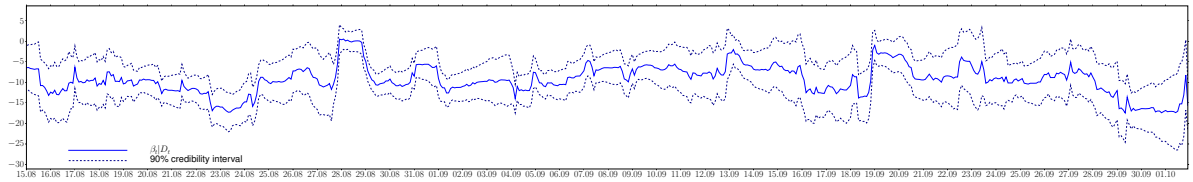
(b) Level



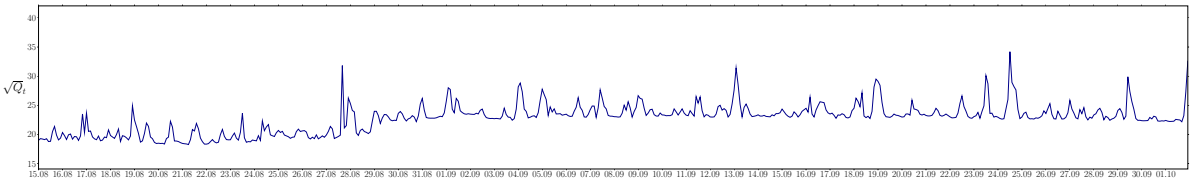
(c) Seasonal effects



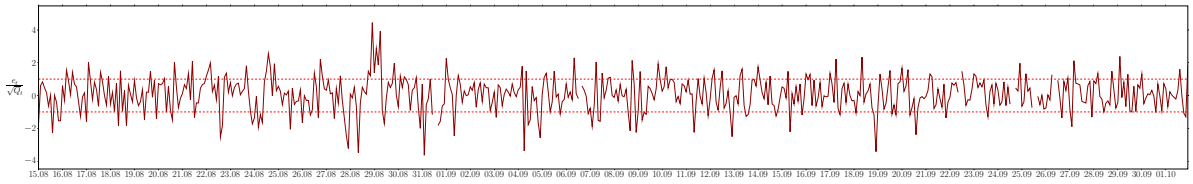
(d) Effect of temperature



(e) Regression coefficient temperature



(f) Prediction standard deviation



(g) Standardized errors

Figure 5.7: M4, D3, $\delta_T = 0.93$, $\delta_S = 1.0$, $\delta_T = 0.91$

For Models 6a and 6b we use statically shifted temperatures

$$\boldsymbol{\theta}_{R_t} = (\beta_{1,t}, \beta_{2,t})'$$
$$\mathbf{F}_{R_t} = (TMP_t^{\#}, TD_t)'$$
$$\mathbf{G}_{R_t} = \mathbf{G}_R = \boldsymbol{I}_2$$

Finally for Models 7a and 7b the temperatures are shifted dynamically

$$\boldsymbol{\theta}_{R_t} = (\beta_{1,t}, \beta_{2,t})'$$
$$\mathbf{F}_{R_t} = (TMP_t^{*}, TD_t)'$$
$$\mathbf{G}_{R_t} = \mathbf{G}_R = \boldsymbol{I}_2$$

It is not necessary to shift the temperature difference, as it over time approximately sums up to zero. The results for all models are given in Table 5.5

Dataset	Model	Burn-in	RMSE	MAD	LLH/T	δ_T	δ_S	δ_R
D1	M5a	48	24.78	18.15	-6.95	0.90	0.98	0.91
D1	M5b	48	24.93	18.52	-6.96	0.89	-	0.90
D1	M6a	48	25.13	18.31	-6.90	0.95	1.0	0.93
D1	M6b	48	25.32	18.55	-6.90	0.95	-	0.93
D1	M7a	48	25.17	18.65	-6.91	0.94	1.0	0.92
D1	M7b	48	25.39	18.73	-6.92	0.94	-	0.92
D2	M5a	48	25.04	18.47	-7.04	0.93	1.0	0.93
D2	M5b	48	25.09	18.71	-7.05	0.94	-	0.94
D2	M6a	48	25.02	18.54	-7.02	0.94	1.0	0.94
D2	M6b	48	25.20	18.65	-7.03	0.94	-	0.92
D2	M7a	48	24.99	18.68	-7.02	0.94	1.0	0.93
D2	M7b	48	25.10	18.73	-7.03	0.95	-	0.93
D3	M5a	48	23.90	17.99	-7.02	0.94	1.0	0.94
D3	M5b	48	23.85	18.09	-7.03	0.95	-	0.95
D3	M6a	48	23.91	18.01	-7.00	0.94	1.0	0.94
D3	M6b	48	23.81	18.09	-7.02	0.95	-	0.96
D3	M7a	48	23.92	18.11	-7.01	0.95	1.0	0.94
D3	M7b	48	23.81	18.09	-7.02	0.96	-	0.96

Table 5.5: Results criteria for Models 5a, 5b, 6a, 6b, 7a and 7b.

The discount factors for all models are quite stable. For Models 5a and 5b the discount factors increase as the dataset gets larger. For Models 6a and 6b the discount factors are quite stable, except for some variation in the regression discount factor, δ_R, for Model 6b. For Models 7a and 7b the discount factors are stable as well, with some increase in the regression discount factor, δ_R, as the dataset grows. Generally the discount factors are high, which indicates little information loss.

Figure 5.8 shows the performance scores for the six new models, as well as Model 2a, 3 and 4. For the full dataset all new models perform better than earlier models. The RMSE scores for

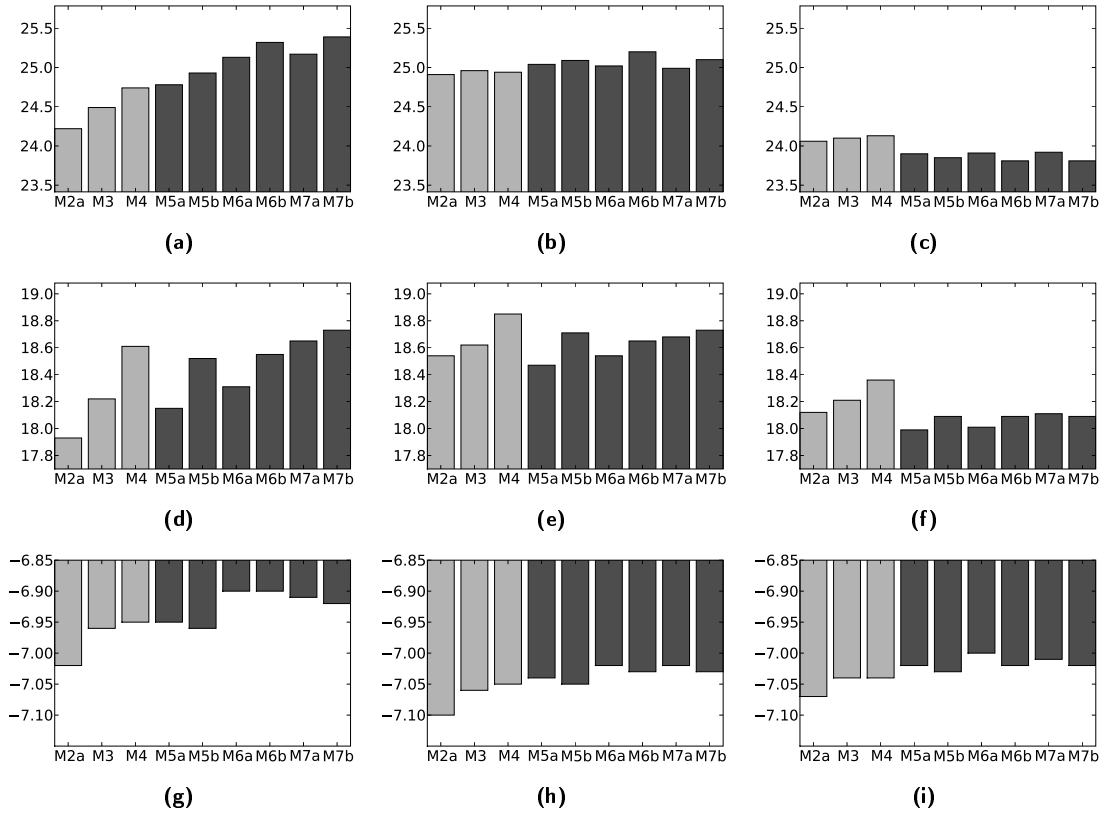
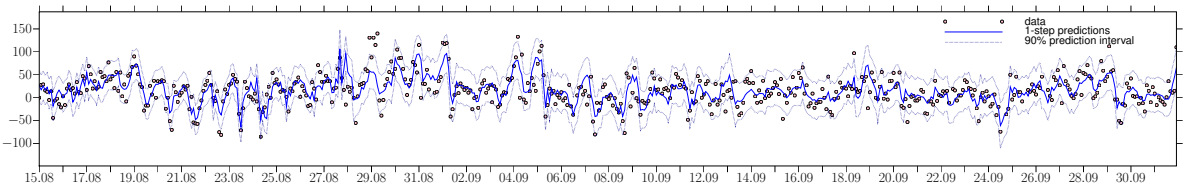


Figure 5.8: Performance scores for models M2a, M3, M4, M5a, M5b, M6a, M6b, M7a and M7b. (a)-(c): RMSE scores for D1, D2 and D3. (d)-(f): MAD scores for D1, D2 and D3. (g)-(i): LLH scores for D1, D2 and D3.

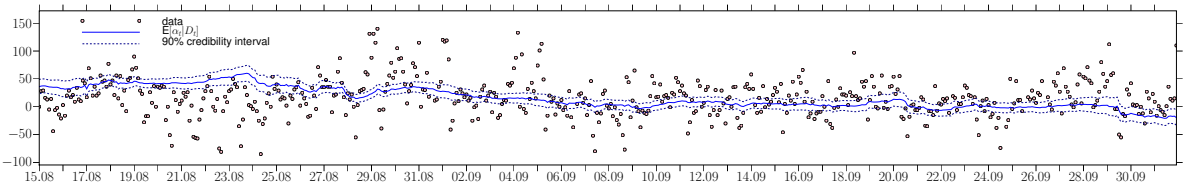
D3 are minimized by Models 6b and 7b. This suggests that the season block is superfluous, and that models with shifted temperatures have better performance. According to the MAD scores, Model 5a is the best model for the full dataset, followed by Model 6a. Thus the non-shifted model is favored, and the season block has a positive effect on the MAD scores for Models 5 and 6, while Model 7b, without the season block, does better than Model 7a. The LLH scores for D3 is maximized for Model 6a. All the season models have better LLH scores than the corresponding models without a season block.

Studying the scores for D1, we see that Model 2a performs better according to RMSE and MAD. It is often the case that competing models have different strengths and weaknesses. As in this case, it can result in different winning models for different parts of a dataset. To utilize the complexity of several models, an alternative is to construct a mixture of several models. Such models are called multi-process models and are described in Chapter 12 in West & Harrison (1997). However, the shifting method of the regression variables must be the same for all models.

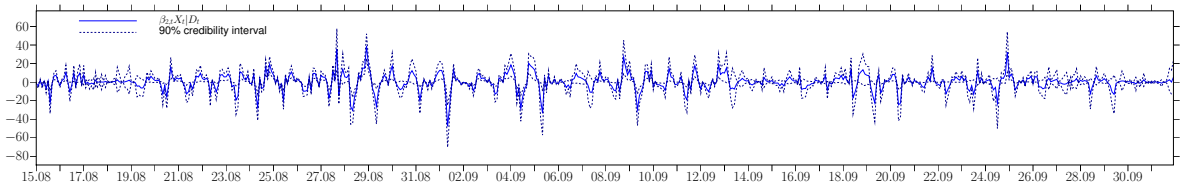
Figure 5.9 shows the results from applying Model 6b with static shifting and no season block, to the full dataset. From Figure 5.9 (a) it is clear that the 1-step prediction still does not reach the more extreme observations. Even though there is improvement in performance compared



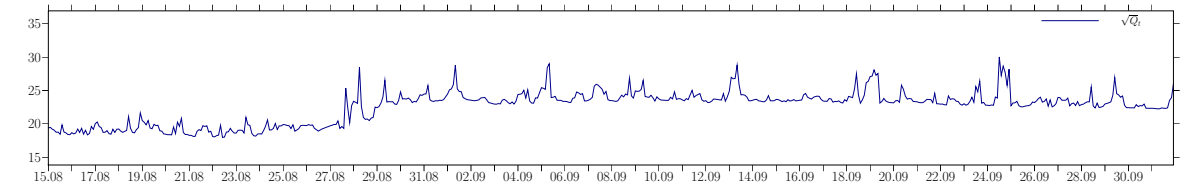
(a) 1-step predictions



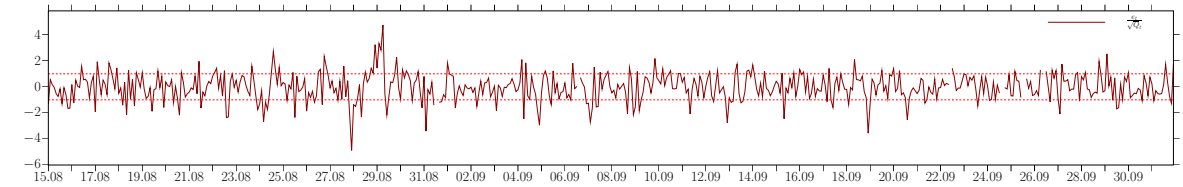
(b) Level



(c) Effect of temperature differences



(d) Prediction standard deviation



(e) Standardized errors

Figure 5.9: M6b, D3, $\delta_T = 0.95$, $\delta_T = 0.96$

to earlier models, adding the temperature difference did not have the desired effect. However Figure 5.9 (b) shows a more stable level with less uncertainty than for earlier models. From Figure 5.9 (c) we can see that the effects of the temperature difference are small, and for the larger part of the dataset they are not statistically significant. From Figure 5.9 (d) we observe that the prediction standard deviation is lower than for earlier models. Figure 5.9 (e) shows the standardized errors, and apparently the problem with correlation in errors is still present.

5.2.5. Model L1, L2 and L3: Form-free Transfer Function. A form-free transfer function, as presented in West & Harrison (1997) Section 9.3.1, is a regression model based on current and past explanatory variables. A fixed number of lagged variables are used in the regression block, hence $Y_{R_t} = \sum_{i=1}^{k} \beta_{i,t}\mathbf{X}_{t-i+1} = \beta_{1,t}X_t + \beta_{2,t}X_{t-1} + ... + \beta_{k,t}X_{t-k}$. The DLM

regression block for a form-free transfer function can be written as

$$\begin{aligned} \boldsymbol{\theta}_{R_t} &= (\beta_{1,t}, \beta_{2,t}, ..., \beta_{k,t})' \\ \mathbf{F}_{R_t} &= (X_t, X_{t-1}, ..., X_{t-k})' \\ \mathbf{G}_{R_t} &= \mathbf{G}_R = \boldsymbol{I}_k \end{aligned} \tag{5.3}$$

If a lagged structure exists and is not modelled, it might explain consecutive, correlated errors as seen for all earlier models.

Models L1a, L2a and L3a are defined as follows:

$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} is a constant trend as defined in (5.2) and Y_{S_t} is single harmonic with period 12 as defined in (5.1). The regression block is a form free-transfer function:

$$\begin{aligned} \boldsymbol{\theta}_{R_t} &= (\beta_{1,t}, \beta_{2,t})' \\ \mathbf{F}_{R_t} &= (X_t, X_{t-2})' \\ \mathbf{G}_{R_t} &= \mathbf{G}_R = \boldsymbol{I}_2 \end{aligned}$$

where X_t is equal to $TMP_t, TMP_t^{\#}$ and TMP_t^* respectively for Models L1a, L2a and L3a. (Notice that we do not regress on X_{t-1}. The model can be written as in (5.3), with $k = 2$ and $\beta_{2,t} = 0$.) Models L1b, L2b and L3b are defined as Models L1a, L2a and L3a respectively, only without the season block:

$$Y_t = Y_{T_t} + Y_{R_t}$$

The results using the six new models are given in Table 5.6. All discount factors are high, which indicates little information loss for all blocks. The regression discount factor, δ_R, is in general higher than for earlier models, which implies more stable regression coefficients. As seen for earlier models, the discount factors increase from D1 to D3, with a some exceptions.

Figure 5.10 shows the predictive performances according to RMSE, MAD and LLH for the six lag models, and a collection of earlier models. The new lag models perform considerably better than earlier alternatives according to all criteria. The lag models without a season block, Models L1b, L2b and L3b, in general give better scores than the season models L1a, L2a and L3a, with some exceptions. Which model performs best varies with dataset and criterion used. For the D3, Model L2b has the best RMSE and LLH scores, while Model L3b minimizes the MAD. Thus the model alternatives without a season block and with shifting is preferred. As experienced for all models, the predictions are better for the first and last part of the full dataset.

Figure 5.11 shows the results applying Model L2b to D3. From Figure 5.11 (a) we see that neither Model L2b has adequate 1-step predictions for the strain values between 27.08 and 29.08, but in general the fit is better than for earlier models. From Figure 5.11 (b), we can see that the level

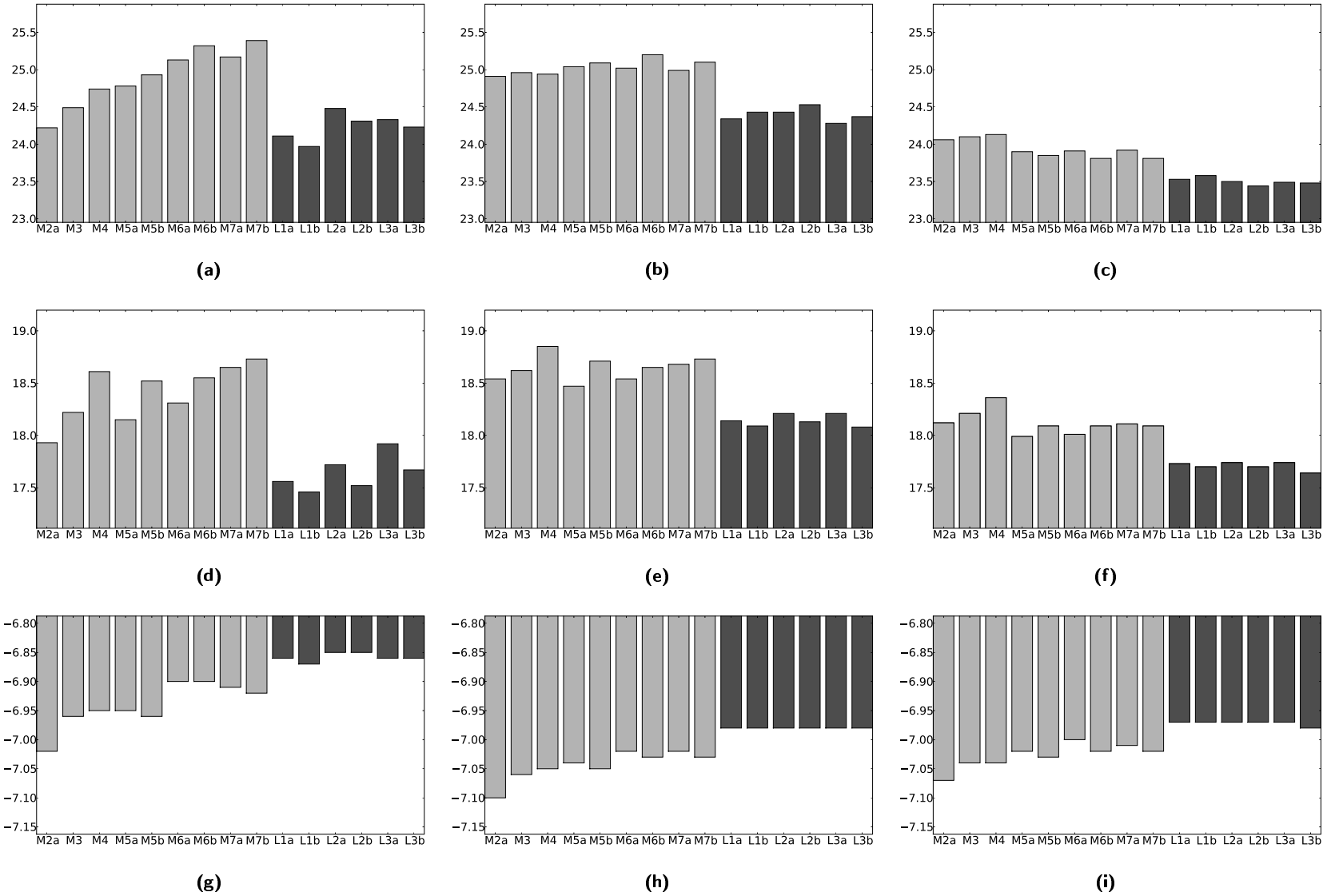


Figure 5.10: Performance scores for models M2a, M3, M4, M5a, M5b, M6a, M6b, M7a, M7b, L1a, L1b, L2a, L2b, L3a and L3b. (a)-(c): RMSE scores for D1, D2 and D3. (d)-(f): MAD scores for D1, D2 and D3. (g)-(i): LLH scores for D1, D2 and D3.

Dataset	Model	RMSE	MAD	LLH	δ_T	δ_S	δ_R
D1	L1a	24.11	17.56	-6.86	0.95	0.99	0.95
D1	L1b	23.97	17.46	-6.87	0.93	-	0.93
D1	L2a	24.48	17.72	-6.85	0.96	1.0	0.95
D1	L2b	24.31	17.52	-6.85	0.95	-	0.93
D1	L3a	24.33	17.92	-6.86	0.95	0.99	0.95
D1	L3b	24.23	17.67	-6.86	0.93	-	0.94
D2	L1a	24.34	18.14	-6.98	0.95	1.0	0.95
D2	L1b	24.43	18.09	-6.98	0.94	-	0.94
D2	L2a	24.43	18.21	-6.98	0.94	1.0	0.96
D2	L2b	24.53	18.13	-6.98	0.94	-	0.94
D2	L3a	24.28	18.21	-6.98	0.96	0.99	0.97
D2	L3b	24.37	18.08	-6.98	0.93	-	0.95
D3	L1a	23.53	17.73	-6.97	0.95	1.0	0.95
D3	L1b	23.58	17.70	-6.97	0.94	-	0.94
D3	L2a	23.50	17.74	-6.97	0.95	1.0	0.96
D3	L2b	23.44	17.70	-6.97	0.95	-	0.96
D3	L3a	23.49	17.74	-6.97	0.96	0.99	0.97
D3	L3b	23.48	17.64	-6.98	0.95	-	0.97

Table 5.6: Results for Models L1a, L2b and L3b.

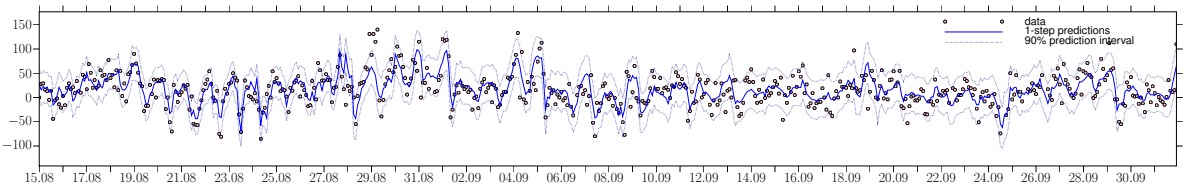
is more stable than for earlier models with static shifting, but there are some noticeable drops in the level (23.08, 27.08, 21.09 and 29.09). As discussed earlier, the level is slowly decreasing over time, caused by the temperature decrease from August to October.

Figure 5.11 (c) shows that the prediction standard deviation is on the same level as Model 6b. From Figure 5.11 (d) we can see that the correlation in the errors are still present. Figures 5.11 (e) and (f) show the posterior regression coefficients $\beta_{1,t}|D_t$ and $\beta_{2,t}|D_t$. In general the coefficients are more stable and has less uncertainty than for earlier models.

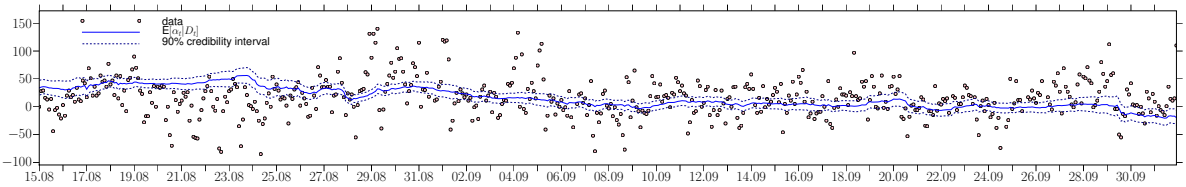
5.2.6. Model F1: Functional Form Transfer Function. For a form-free transfer function as described in the previous section, X_t influences the response variable for only a limited time-span. X_t ceases to have any effect on Y_{t+r} when $r > k$. Hence, X_t can in principle have major influence on Y_{t+k}, but has no effect at all on Y_{t+k+1}. For a functional form transfer function, as described in West & Harrison (1997) Section 9.3.2, the effect of a regression variable decreases smoothly over time, gradually tending to zero.

The regression block for a one dimensional functional form transfer function can be written as

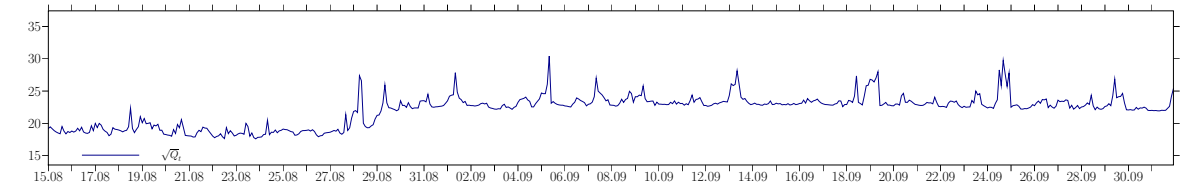
$$\begin{aligned} Y_{R_t} &= \xi_t \\ \xi_t &= \lambda\xi_{t-1} + \psi_t X_t + \partial\xi_t \\ \psi_t &= \psi_{t-1} + \partial\psi_t, \end{aligned} \tag{5.4}$$



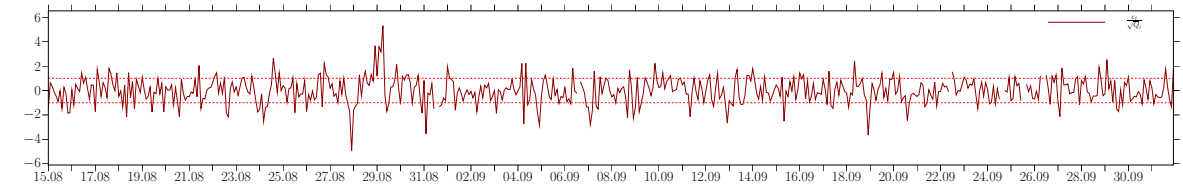
(a) 1-step predictions



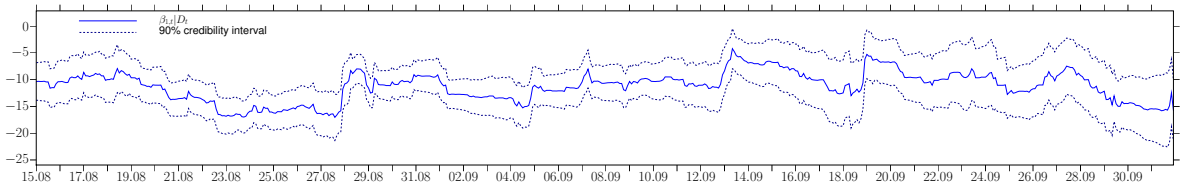
(b) Level



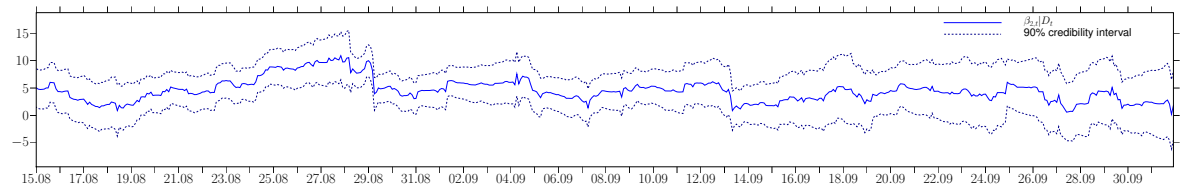
(c) Prediction standard deviation



(d) Standardized errors



(e) $\beta_{1,t}|D_t$



(f) $\beta_{2,t}|D_t$

Figure 5.11: L2b, D3, $\delta_T = 0.95$, $\delta_R = 0.96$

where ψ_t represents the *penetration* effect, which is the immediate effect of the present regression variable X_t. λ denotes the *memory* and represents the effect of past regressors. Let $h_t = E[\psi_t|D_t]$. Then a k-step forecast function for the regression block is given by

$$f_{R_t}(k) = \lambda^k \xi_t + h_t \sum_{i=1}^{k} \lambda^{k-r} X_{t+r}$$

Thus if $\lambda \in (0,1)$, the effect of X_t will be reduced by the factor λ for each point of time, and diminish as k grows.

The functional form transfer function can be written in the general DLM form defined in (2.1). Using the superpositioned DLM in (2.3), the trend and season block are defined as for regular DLMs. (5.4) can be expressed as a regression block as follows:

$$\begin{aligned} \boldsymbol{\theta}_{\boldsymbol{R_t}} &= (\xi_t, \psi_t)' \\ \mathbf{F}_{R_t} &= \mathbf{F_R} = (1,0)' \\ \mathbf{G}_{R_t} &= \begin{pmatrix} \lambda & X_t \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{5.5}$$

Furthermore, let $U_t = \text{Var}[\partial\xi_t]$ and $Z_t = \text{Var}[\partial\psi_t]$. If $\partial\xi_t$ and $\partial\psi_t$ are assumed to be uncorrelated (which is most direct and appropriate alternative according to West & Harrison (1997)), then

$$\mathbf{W}_{R_t} = \begin{pmatrix} U_t + {X_t}^2 Z_t & X_t Z_t \\ X_t Z_t & Z_t \end{pmatrix}$$

U_t and Z_t are most conveniently dealt with using discount factors. Let

$$\mathbf{P}_{R_t} = \mathbf{G}_{R_t}\mathbf{C}_{R_{t-1}}\mathbf{G}_{R_t}$$

Then it follows that

$$\begin{aligned} U_t &= \frac{1-\delta_R}{\delta_R}\{\mathbf{P}_{R_t}\}_{1,1} \\ Z_t &= \frac{1-\delta_R}{\delta_R}\{\mathbf{P}_{R_t}\}_{2,2} \end{aligned}$$

The regression block described includes only one explanatory variable, but can of course be extended to include multiple variables, either as transfer functions or other types of regression functions. A general functional form transfer function is defined in West & Harrison (1997) page 284 as

$$\begin{aligned} Y_t &= \tilde{\mathbf{F}}'\tilde{\boldsymbol{\theta}}_t + \nu_t \\ \tilde{\boldsymbol{\theta}}_t &= \tilde{\mathbf{G}}\tilde{\boldsymbol{\theta}}_{t-1} + \psi_t X_t + \partial\tilde{\boldsymbol{\theta}}_t \\ \psi_t &= \psi_{t-1} + \partial\psi_t, \end{aligned}$$

where $\tilde{\mathbf{F}}$ is a known vector, $\tilde{\boldsymbol{\theta}}_t$ is the unknown state vector and $\tilde{\mathbf{G}}$ is a constant, known evolution matrix. ν_t is the observation error and $\partial\tilde{\boldsymbol{\theta}}_t$ is the evolution error, assumed to be mutually independent. X_t is the current regressor and ψ_t is a parameter modelled as a simple random walk, with a noise term $\partial\psi_t$ which is assumed to be independent of ν_t and normally distributed with a zero mean. $\partial\tilde{\boldsymbol{\theta}}_t$ and $\partial\boldsymbol{\psi}_t$ are not necessarily independent of each other.

Model F1a has the familiar block structure

$$Y_t = Y_{T_t} + Y_{S_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} and Y_{S_t} are as defined in (5.2) and (5.1) respectively. Y_{R_t} is as specified in (5.5). For the lag models temperature was used as the explanatory variable. For this model however, temperature differences give the best results, so $X_t = TD_t$. After exploring several alternatives, the memory parameter λ is set to 0.78. An alternative to a deterministic λ is to include a learning procedure on the parameter, and let λ_t be stochastic. This is discussed in Section 9.3.3 in West & Harrison (1997).

Model F1b is defined as

$$Y_t = Y_{T_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where the trend and regression blocks are as defined for Model F1a. Table 5.7 shows the results for the two models. The discount factors for all blocks are quite stable. The trend discounts,

Dataset	Model	RMSE	MAD	LLH	δ_T	δ_S	δ_R
D1	F1a	23.95	18.08	-6.87	0.80	0.98	0.97
D1	F1b	24.21	18.32	-6.91	0.78	-	0.97
D2	F1a	24.64	18.82	-6.99	0.81	0.99	0.97
D2	F1b	24.81	18.97	-7.03	0.80	-	0.97
D3	F1a	24.10	18.52	-6.99	0.80	0.99	0.98
D3	F1b	24.14	18.51	-7.02	0.79	-	0.98

Table 5.7: Results for Models F1a and F1b.

varying between 0.80 and 0.81 for Model F1a, and 0.78 to 0.80 for Model F1b are rather low, which might imply that the model is insufficient.

From Figure 5.12 we can see that Model F1a with a season block performs better than Model F1b. The MAD scores for D3 is the only exception, but the differences between the two models are marginal. None of the transfer models can compete with the lag models. However, Model F1a has a very good RMSE score for D1.

Figure 5.13 shows the results from applying Model F1a to D3. From the 1-step prediction in Figure 5.13 (a), we can see that the model performs quite well for the extreme observations on the 27.08 and 29.08, compared to Model L2b. The explanation can be found in Figure 5.13 (b). Since the trend block has a low discount factor, the level is allowed to change rapidly. Thus the 1-step prediction adjusts much faster to extreme behavior. This is beneficial for the period in question, but in general it leads to high uncertainty in the level and is a result of insufficient modeling. The trend discount factor for the lag models were very high, and thus the trend can not adjust to sudden changes. Allowing lower discount factors for some periods might improve the model performance.

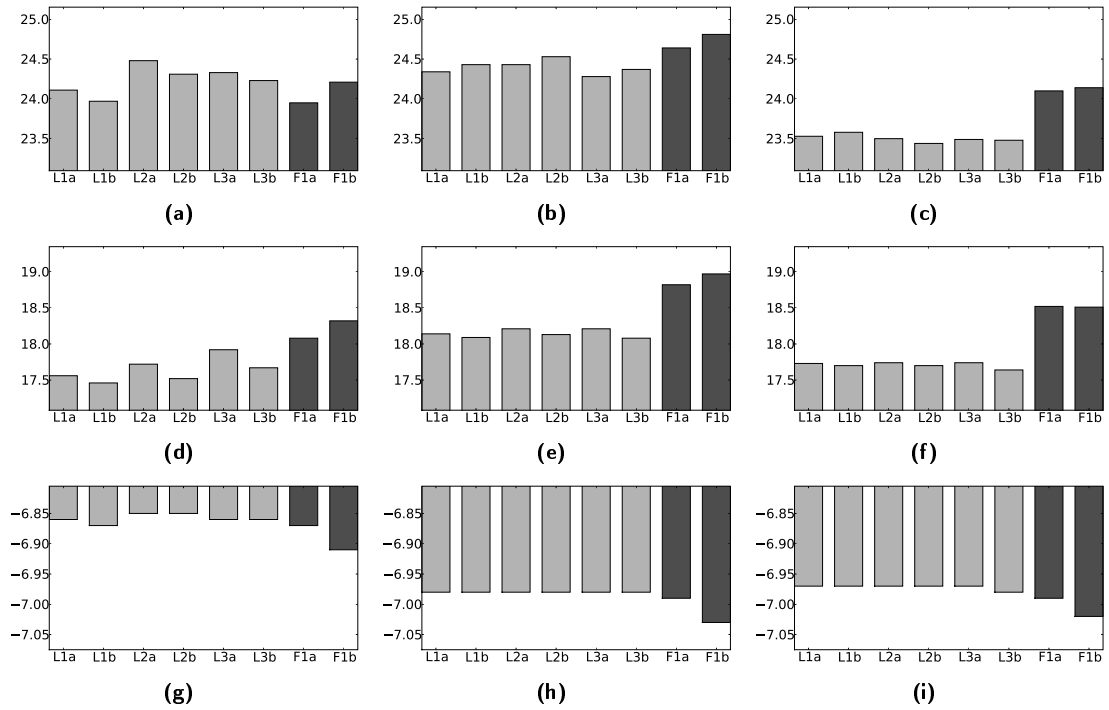
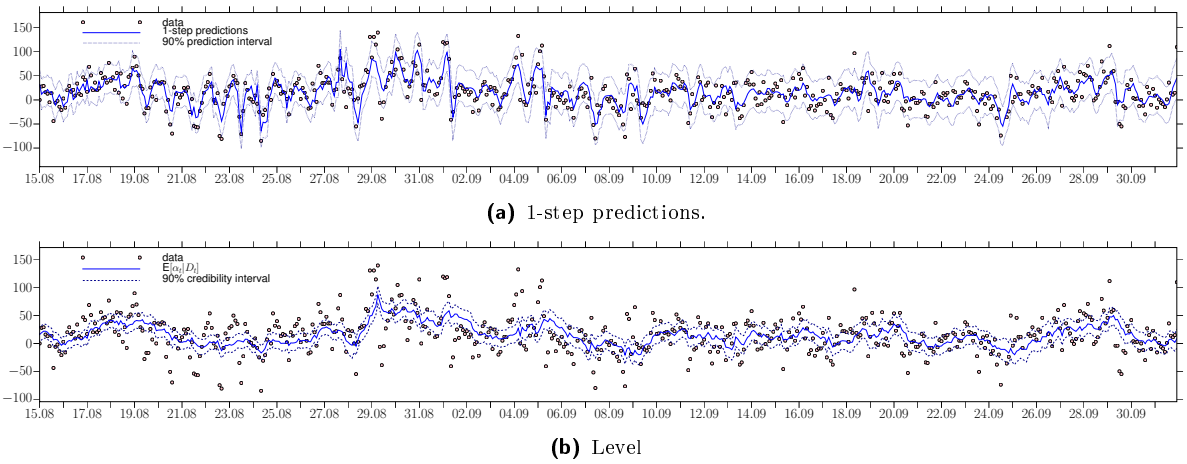


Figure 5.12: Performance scores for models L1a, L1b, L2a, L2b, L3a, L3b, F1a and F1b. (a)-(c): RMSE scores for D1, D2 and D3. (d)-(f): MAD scores for D1, D2 and D3. (g)-(i): LLH scores for D1, D2 and D3.



(a) 1-step predictions.

(b) Level

Figure 5.13: F1a, D3, $\delta_T = 0.80$, $\delta_S = 0.99$ $\delta_R = 0.98$

5.2.7. Model R1: Radiation. As suggested in Chapter 3, we will test radiation as a regression variable. So far, Model L2b has the best overall performance of all suggested models, thus we use L2b as a basis. Model R1 is defined as

$$Y_t = Y_{T_t} + Y_{R_t} + \nu_t, \qquad \nu_t \sim N[0, V],$$

where Y_{T_t} is defined in (5.2), and

$$\boldsymbol{\theta}_{R_t} = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$$
$$\mathbf{F}_{R_t} = (TMP_t^{\#}, TMP_{t-2}^{\#}, RAD_t)'$$
$$\mathbf{G}_{R_t} = \mathbf{G}_R = \boldsymbol{I}_3$$

RAD_t is the global radiation at time t, shifted so that the average contribution from global radiation is equal to zero. The performance scores for Model R1 are given in Table 5.8. Except

Dataset	Model	RMSE	MAD	LLH	δ_T	δ_R
D1	R1	24.54	17.74	-6.82	0.96	0.95
D2	R1	24.34	18.14	-6.97	0.94	0.95
D3	R1	24.32	17.67	-6.80	0.95	0.95

Table 5.8: Results for Model R1.

for the RMSE and MAD scores for D2, all scores imply that Model L2b performs better or equal to Model R1. Figure 5.14 shows the effect from radiation. The effect is most of the time very close to zero, and is seldom significant at a 10% level. Hence radiation is excluded as an explanatory variable.

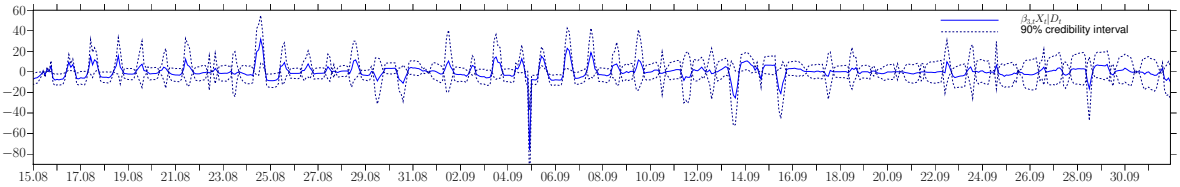


Figure 5.14: R1a, D3, $\delta_T = 0.95$, $\delta_R = 0.95$. Effect of radiation.

5.3. Forecasting with Regression Models

Forecasting with regression models requires knowledge about future explanatory variables. In most cases, as for temperatures in our application, these are not known. This introduces an extra uncertainty to the model. There are several ways of handling unknown regression variables. One alternative is to use a joint forecasting model for Y_t and the regression variable $\mathbf{X}_t$, where $\mathbf{X}_t$ is modelled as a response variable in a separate time series. Such multivariate time series models are described in Chapter 16, West & Harrison (1997).

A simpler method is to use estimated densities for future regression variables. The densities can be separate models or simply qualified guesses. With a known density function for the regression vector, $p(\mathbf{F}_{t+k}|D_t)$, the step ahead prediction is given by

$$p(Y_{t+k}|D_t) = \int p(Y_{t+k}|\mathbf{F}_{t+k}, D_t) p(\mathbf{F}_{t+k}|D_t) d\mathbf{F}_{t+k}$$

The fully specified forecast distribution depends on the distribution of the future regression vector. The step ahead mean and variation of Y_{t+k} can also be deduced only knowing $\mathrm{E}[\mathbf{F}_{t+k}|D_t]$ and $\mathrm{Var}[\mathbf{F}_{t+k}|D_t]$. For more details see Section 9.2.4 in (West & Harrison, 1997).

When no knowledge on future regression variables is available, a solution is to do what West & Harrison (1997) refers to as 'What if?' predictions. Such predictions are based on a set of hypothetical future regressors. This way we can study forecasts given different scenarios. There are few phenomena that are as well covered as weather. Temperature forecasts are available at a range of websites. Hourly forecasts for Gothenburg can for example be found at www.yr.no.

Deciding which external factors influence the response is crucial when building a model. The step ahead predictions in this chapter are therefore based on foresight. The foresight is limited only to the regression variables. (Foresight in the response variable would be equal to doing a retrospective analysis). Real-time forecasting requires use of the methods discussed in this section.

5.4. Conclusions

Temperature was found to be an important explanatory factor for the strain response. In addition to using current temperature as an explanatory variable, also regressing on lagged temperatures, by using transfer models, improved the performance. Form-free transfer functions gave better results than functional form transfer functions, and it was found to be optimal to regress on X_t and X_{t-2}, where X_t represents temperatures at time t. Using temperatures which were shifted statically gave the best 1-step predictions, while the level parameter with dynamic shifted temperatures gave a better understanding of the data's real underlying level. The season block was found to be redundant for these lag models, and all in all Model L2b had the best performance of all suggested models. Adding radiation as an explanatory variable did not improve the model. Even though Model L2b has overall good predictions, the performance deteriorates for certain periods. For all models, the predictions were best for the first and latter part of the dataset. The optimal discount factors are quite stable throughout the dataset, but allowing lower discount factors for some periods might improve the model. We experienced that Model 1a and Model F1a, which both had low trend discount factors, had smaller errors for some of the more extreme observations.

CHAPTER 6

Step Ahead Predictions

In Chapter 5 we compared a number of models on the basis of their 1-step ahead forecast performance. In this chapter we will take a closer look at some of the models, and examine the quality of their k-step forecasts. The models selected are Models 1a, 2a, L1b, L2b and L3b. The three latter, all described in detail in Section 5.2.5, is a natural choice since these models had the best performance scores for 1-step ahead forecasts. Model L2b had the overall best performance, but L1b and L3b had good results as well. It could be interesting to see if the different shifting methods have any effect on the step ahead forecasts. Model 2a, defined in Section 5.2.2, is the simplest of the regression models, with air temperature as the only regression variable. Thus the forecast process is less complex if future temperatures are unknown, as is the case when doing on-line forecasting. Finally Model 1a, defined in Section 5.2.1, will be tested since it has no regression block and therefore does not require any external information. It is also interesting to see how robust a simple trend and season model is compared to more complex regression models.

For each model we perform k-step prediction with starting point at three different times: 30.08 at 22.00 Hrs, 15.09 at 22.00 Hrs and 01.10 at 22.00 Hrs. Thus the step ahead predictions are based on all information available the given date. As for the 1-step predictions, future temperature values are assumed known. The selected dates represent the end points of the data fractions D1, D2 and D3. The discount factors used are listed in Table 6.1, and are equal to the optimal factors found in Chapter 5, one set of values for each data fraction. To examine both short and long term prediction performance, a range of k-values are tested: 6, 12, 24, 36, 48, 60 and 120. With 12 observations per day, these are equal to 0.5, 1, 2, 3, 4, 5 and 10 days respectively.

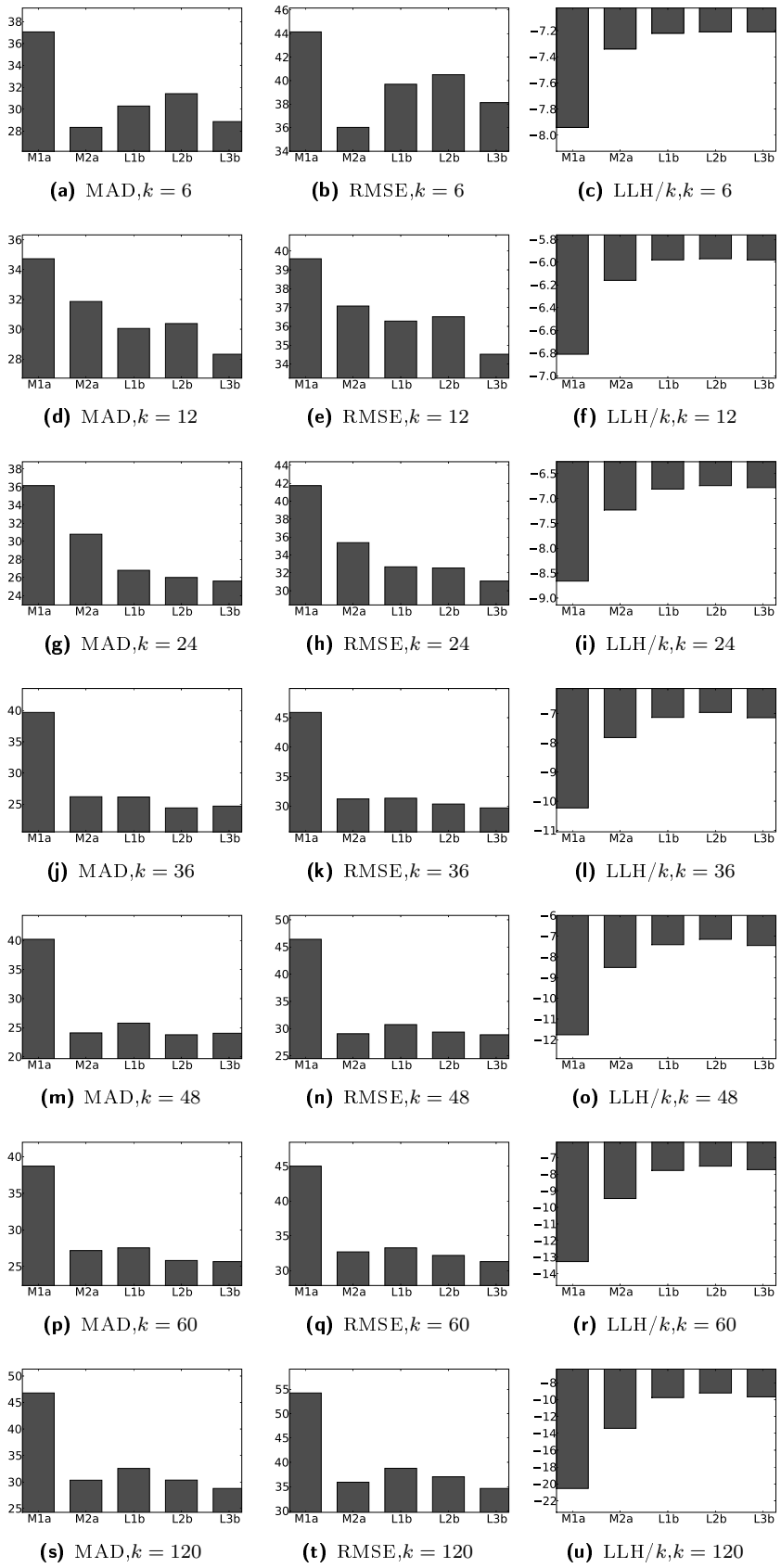
Model	Dataset	δ_T	δ_S	δ_R
M1a	D1	0.78	-	0.88
M1a	D2	0.80	-	0.94
M1a	D3	0.77	-	0.98
M2a	D1	0.88	0.97	0.89
M2a	D2	0.90	0.99	0.91
M2a	D3	0.91	0.99	0.92
L1b	D1	0.93	-	0.93
L1b	D2	0.94	-	0.94
L1b	D3	0.94	-	0.94
L2b	D1	0.95	-	0.93
L2b	D2	0.94	-	0.94
L2b	D3	0.95	-	0.96
L3b	D1	0.93	-	0.94
L3b	D2	0.93	-	0.95
L3b	D3	0.95	-	0.97

Table 6.1: Discount factors used for k-step predictions.

6.1. Results for Prediction made 30.08 at 22 Hrs

The results for k-step predictions with starting point 30.08 at 22 Hrs, for Models 1a, 2a, L1b, L2b and L3b are given in Table 6.2. Figure 6.1 shows the performance scores for all models and k-values. The results clearly show that Models L2b and L3b, the lag models with static and dynamic shifting, has the overall best k-step predictions according to all three criteria. However, there are some exceptions. The simpler Model 2a, with non-shifted temperature regression, trend and a season block, is the best model according to MAD and RMSE for $k = 6$. For larger k-values Models L2b and L3b are far better, but Model 2a partly competes with the lag model L1b. Model 2a does however have poorer LLH/k scores for all k-values. In Chapter 5, we saw that the uncertainties of some of the parameters in Model 2a were high, and since the k-step predictions give additive uncertainty, that accounts for the lower LLH/k scores and explains why the scores deteriorate with the k-value. The plain trend and season Model 1a has the poorest performance scores of the models according to all criteria. The LLH/k scores for Model 1a are inferior compared to all other models, especially for large k-values.

For all models the LLH/k scores decrease with k, with one exception from $k = 6$ to $k = 12$. The MAD and RMSE scores are quite stable, and for Model 2a, L1b, L2b and L3b, $k = 48$ gives the best RMSE and MAD scores. Since MAD $= \sum_{i=t}^{t+k} |e_i|/k$ and RMSE $= \sqrt{\sum_{i=t}^{t+k} e_i^2/k}$, both measures, as well as LLH/k, are more sensitive to outliers for small k values. Thus one outlier in the first 6 predictions might explain the improvement we see in the scores from $k = 6$ to $k = 12$. For k-values greater than 48, the MAD and RMSE scores increase, and from $k = 60$ to $k = 120$, there is considerable jump. Thus the point predictions do deteriorate over time also according to MAD and RMSE, but rather slowly.



(a) MAD,$k = 6$ **(b)** RMSE,$k = 6$ **(c)** LLH/k,$k = 6$

(d) MAD,$k = 12$ **(e)** RMSE,$k = 12$ **(f)** LLH/k,$k = 12$

(g) MAD,$k = 24$ **(h)** RMSE,$k = 24$ **(i)** LLH/k,$k = 24$

(j) MAD,$k = 36$ **(k)** RMSE,$k = 36$ **(l)** LLH/k,$k = 36$

(m) MAD,$k = 48$ **(n)** RMSE,$k = 48$ **(o)** LLH/k,$k = 48$

(p) MAD,$k = 60$ **(q)** RMSE,$k = 60$ **(r)** LLH/k,$k = 60$

(s) MAD,$k = 120$ **(t)** RMSE,$k = 120$ **(u)** LLH/k,$k = 120$

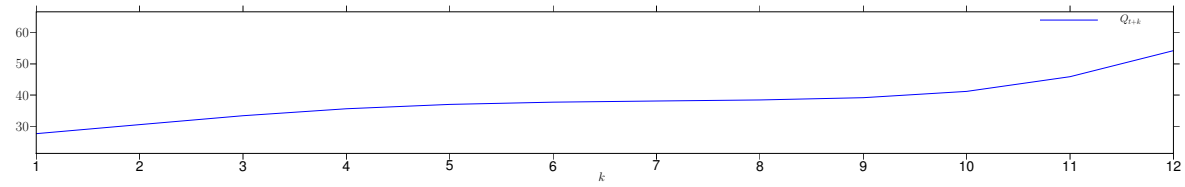
Figure 6.1: Performance scores for step ahead forecasting with Models 1a, 2a, L1b, L2b and L3b. k-values 6, 12, 24, 36, 48, 60 and 120 are used, which corresponds to respectively 1/2, 1, 2, 3, 4, 5 and 10 days. The predictions are done with all information available 30.08 at 22.00 Hrs

Model	k	MAD	RMSE	LLH/k
M1a	6	37.06	44.14	-7.94
M1a	12	34.73	39.59	-6.81
M1a	24	36.16	41.74	-8.66
M1a	36	39.69	45.89	-10.23
M1a	48	40.20	46.41	-11.76
M1a	60	38.71	45.02	-13.27
M1a	120	46.80	54.25	-20.54
M2a	6	28.34	36.02	-7.34
M2a	12	31.86	37.08	-6.16
M2a	24	30.78	35.37	-7.23
M2a	36	26.19	31.21	-7.82
M2a	48	24.10	29.05	-8.51
M2a	60	27.18	32.68	-9.47
M2a	120	30.34	35.89	-13.41
L1b	6	30.29	39.69	-7.22
L1b	12	30.05	36.28	-5.98
L1b	24	26.79	32.67	-6.81
L1b	36	26.15	31.33	-7.13
L1b	48	25.78	30.73	-7.41
L1b	60	27.55	33.26	-7.78
L1b	120	32.56	38.76	-9.77
L2b	6	31.42	40.50	-7.21
L2b	12	30.38	36.51	-5.97
L2b	24	26.00	32.54	-6.74
L2b	36	24.40	30.34	-6.96
L2b	48	23.78	29.35	-7.15
L2b	60	25.82	32.17	-7.52
L2b	120	30.36	37.01	-9.22
L3b	6	28.87	38.13	-7.21
L3b	12	28.32	34.52	-5.98
L3b	24	25.60	31.08	-6.78
L3b	36	24.69	29.68	-7.14
L3b	48	24.04	28.86	-7.45
L3b	60	25.67	31.28	-7.73
L3b	120	28.78	34.61	-9.68

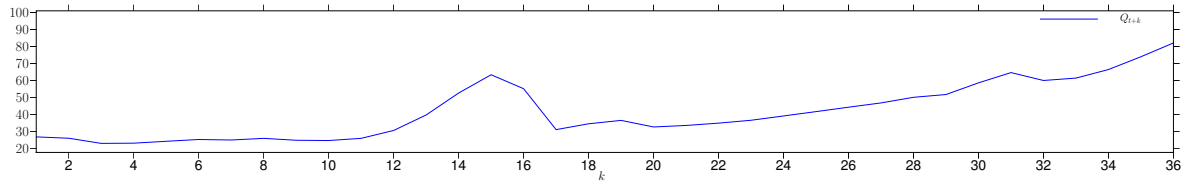
Table 6.2: Performance scores for k-step prediction made 30.08 at 22.00Hrs for Models 1a, 2a, L1b, L2b and L3b.

Figure 6.2 shows the prediction standard deviation for a range of k-values for the five tested models. We see from Figure 6.2 (a) that the standard deviation for Model 1a quickly increases, even for low k-values. Figure 6.2 (b) shows that the prediction standard deviation for Model 2a is stable till $k = 11$, followed by a jump. The prediction standard deviation for Models L1b, L2b and L3b are shown in Figures 6.2 (c), (d) and (e). For all three models the standard deviation slowly increases from around $k = 10$, but is fairly low till around $k = 40$.

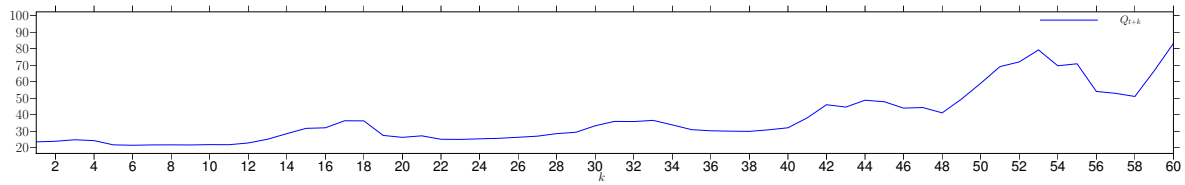
Figure 6.3 shows k-step predictions for Models 1a, 2a, L1b, L2b and L3b. There are two missing values at $k = 7$ and $k = 8$, corresponding to 31.08 at 12.00 Hrs and 14.00 Hrs. For all models,



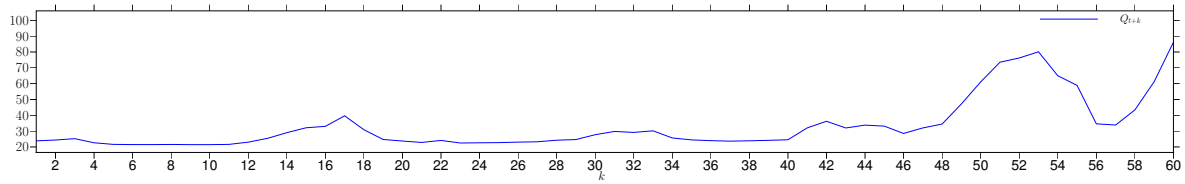
(a) M1a. k-values ranging from 1 to 12



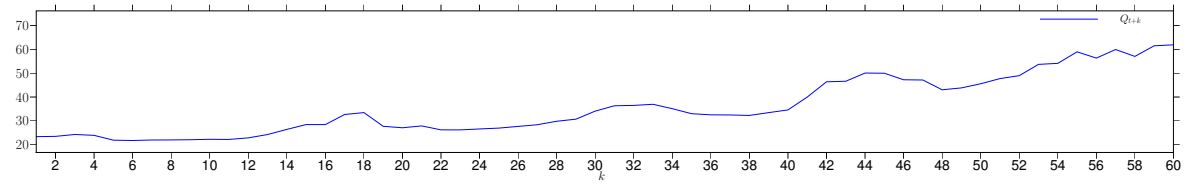
(b) M2a. k-values ranging from 1 to 36



(c) L1b. k-values ranging from 1 to 60



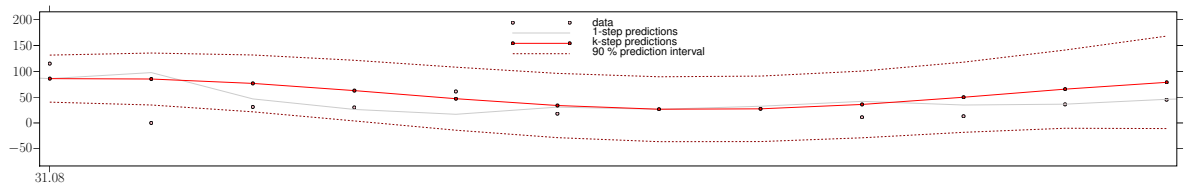
(d) L2b. k-values ranging from 1 to 60



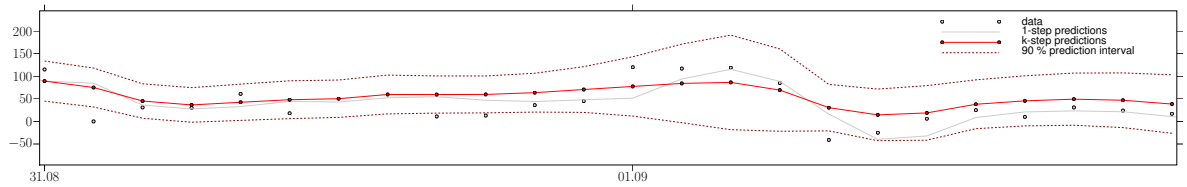
(e) L3b. k-values ranging from 1 to 60

Figure 6.2: Standard deviation for k-step predictions made 30.08 at 22.00 Hrs

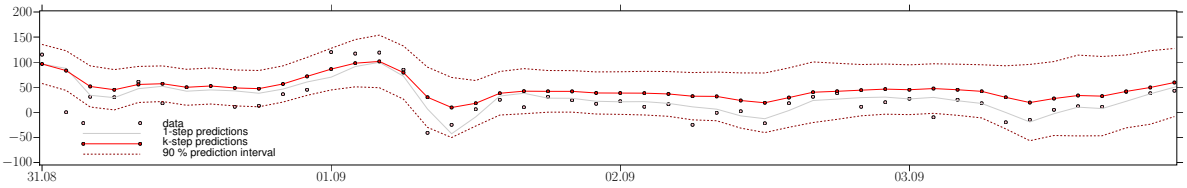
there is an outlier at $k = 2$. This explains why $k = 12$ gave better performance results than $k = 6$, as the outlier has more influence for smaller ks. Model 1a does not have as accurate predictions as the other models and the prediction interval increases rapidly. The predictions for Model 1a are unreliable, even for small ks and it seems as the seasonal pattern does not fit the data motion. Even though Model 2a partly had MAD and RMSE scores competing with the lag models, it is evident that the uncertainties are too large to give the predictions any credibility for k-values larger than 11, where the prediction interval around the point predictions drastically increases. The lag models have far more narrow prediction intervals, especially for larger k-values. The prediction interval for Model L2b is quite stable until $k = 40$, 03.09 at 06.00 Hrs. For Models L1b and L3b, the intervals increase a few steps earlier. For all three models, most of the observations lie within the 90 % interval, with some exceptions. However, from 01.09



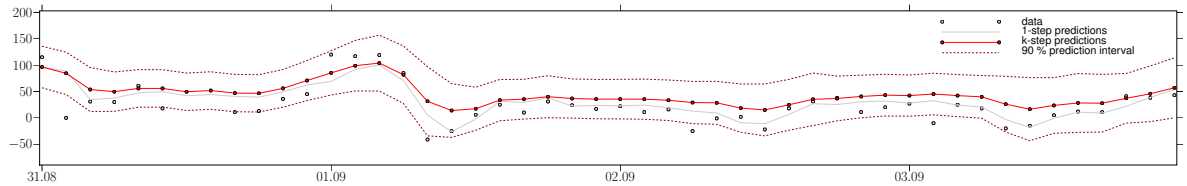
(a) M1a. k-values ranging from 1 to 12



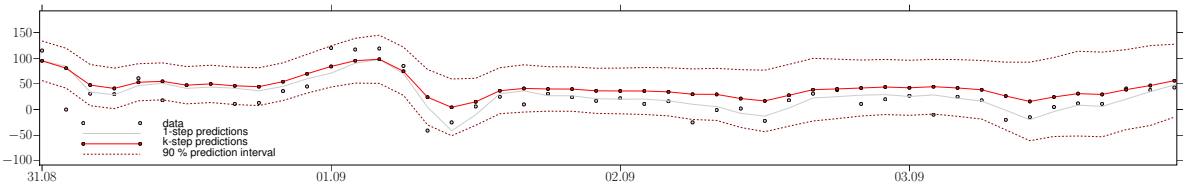
(b) M2a. k-values ranging from 1 to 24



(c) L1b. k-values ranging from 1 to 48



(d) L2b. k-values ranging from 1 to 48



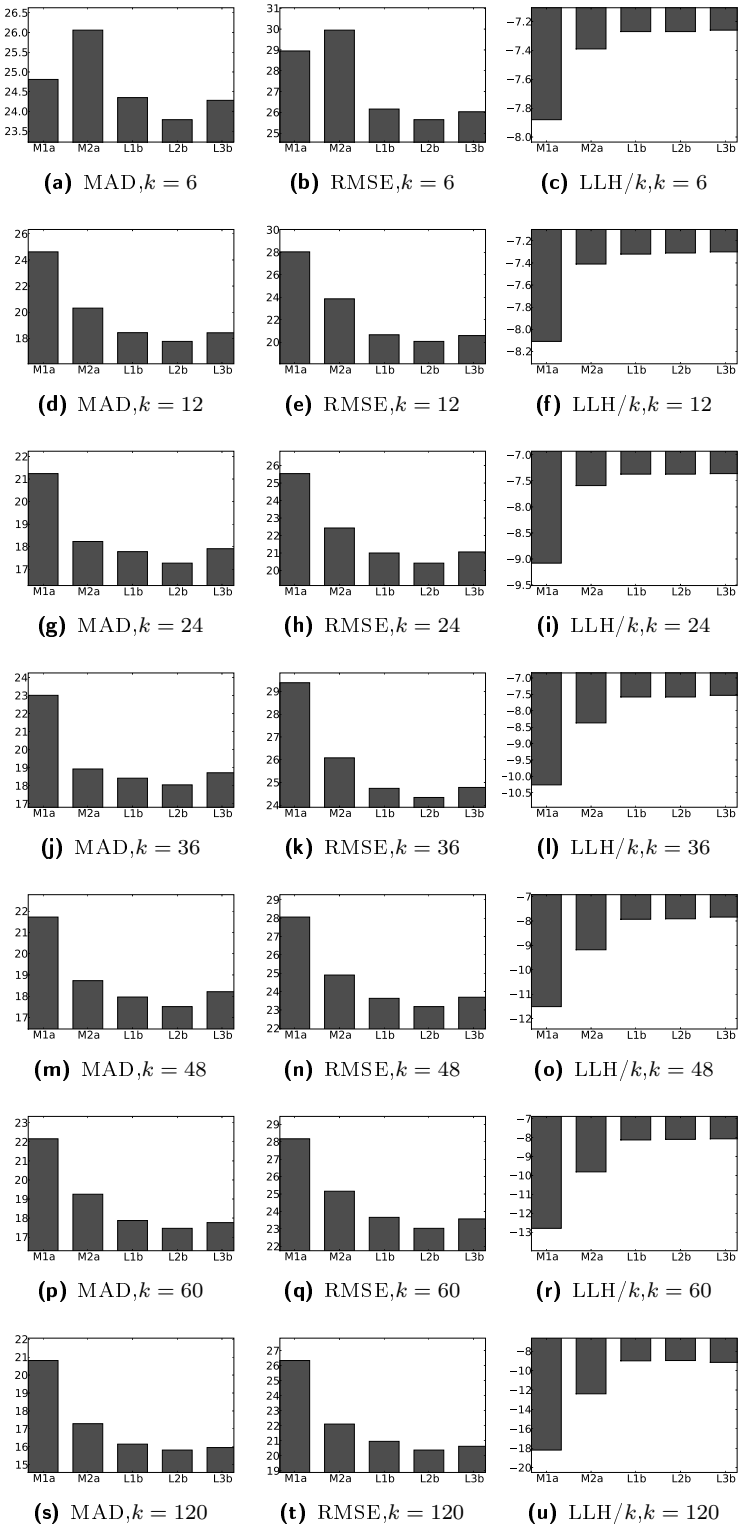
(e) L3b. k-values ranging from 1 to 48

Figure 6.3: k-step predictions made 30.08 at 22.00 Hrs

at 08.00 Hrs corresponding to $k = 16$, we can see that the three lag models consistently predict above the observed values. 1-step predictions on the other hand adjust to the new level, which illustrates the strength of the dynamic parameters in the DLM. We have to take into account that such shifts in parameters may occur, and it is crucial to update k-step predictions each time new data or other information of relevance is available.

6.2. Results for Prediction made 15.09 at 22 Hrs

In Table 6.3, the performance scores for Models 1a, 2a, L1b, L2b and L3b for predictions made 15.09 at 22.00Hrs are listed. In Figure 6.4 the same results are plotted.



(a) MAD,$k = 6$ **(b)** RMSE,$k = 6$ **(c)** LLH/k,$k = 6$

(d) MAD,$k = 12$ **(e)** RMSE,$k = 12$ **(f)** LLH/k,$k = 12$

(g) MAD,$k = 24$ **(h)** RMSE,$k = 24$ **(i)** LLH/k,$k = 24$

(j) MAD,$k = 36$ **(k)** RMSE,$k = 36$ **(l)** LLH/k,$k = 36$

(m) MAD,$k = 48$ **(n)** RMSE,$k = 48$ **(o)** LLH/k,$k = 48$

(p) MAD,$k = 60$ **(q)** RMSE,$k = 60$ **(r)** LLH/k,$k = 60$

(s) MAD,$k = 120$ **(t)** RMSE,$k = 120$ **(u)** LLH/k,$k = 120$

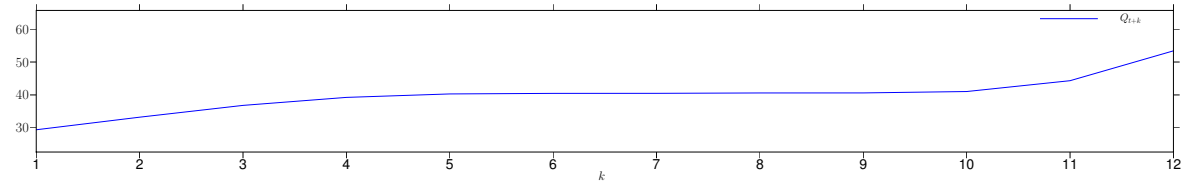
Figure 6.4: Performance scores for step ahead forecasting with Models 1a, 2a, L1b, L2b and L3b. k-values 6, 12, 24, 36, 48, 60 and 120 are used, which corresponds to respectively 1/2, 1, 2, 3, 4, 5 and 10 days. The predictions are done with all information available 15.09 at 22.00 Hrs.

Model	k	MAD	RMSE	LLH/k
M1a	6	24.81	28.95	-7.88
M1a	12	24.62	28.04	-8.11
M1a	24	21.24	25.54	-9.08
M1a	36	23.01	29.38	-10.26
M1a	48	21.72	28.06	-11.51
M1a	60	22.16	28.17	-12.80
M1a	120	20.81	26.34	-18.19
M2a	6	26.06	29.95	-7.39
M2a	12	20.31	23.86	-7.41
M2a	24	18.23	22.43	-7.59
M2a	36	18.93	26.08	-8.37
M2a	48	18.73	24.90	-9.18
M2a	60	19.25	25.16	-9.81
M2a	120	17.29	22.11	-12.39
L1b	6	24.35	26.16	-7.27
L1b	12	18.43	20.66	-7.32
L1b	24	17.78	21.00	-7.37
L1b	36	18.42	24.74	-7.58
L1b	48	17.96	23.63	-7.93
L1b	60	17.87	23.66	-8.12
L1b	120	16.15	20.96	-8.99
L2b	6	23.79	25.65	-7.27
L2b	12	17.76	20.07	-7.31
L2b	24	17.27	20.42	-7.37
L2b	36	18.05	24.34	-7.58
L2b	48	17.51	23.18	-7.91
L2b	60	17.46	23.03	-8.09
L2b	120	15.82	20.38	-8.96
L3b	6	24.28	26.03	-7.26
L3b	12	18.42	20.59	-7.30
L3b	24	17.91	21.06	-7.36
L3b	36	18.72	24.78	-7.53
L3b	48	18.21	23.69	-7.84
L3b	60	17.76	23.57	-8.06
L3b	120	15.96	20.63	-9.15

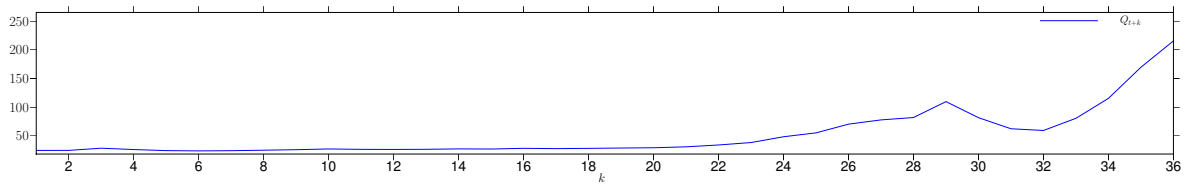
Table 6.3: Performance scores for k-step prediction made 15.09 at 22.00Hrs for Models 1a, 2a, L1b, L2b and L3b.

In general, all models have considerably better prediction performances 15.09 than 30.08 according to MAD and RMSE. In Figure 3.4 (a), we observe that the data are more stable the days succeeding 15.09 than the data in the period after 30.08. This might indicate that our models do better prediction for more stable periods. The LLH/k scores were on the other hand better for the predictions made 30.08. Since the point predictions made 15.09 are more accurate, the poorer LLH/k scores must be related to the uncertainty of the predictions. For all k-values, Model L2b has the best predictions according to the MAD and RMSE scores, while Models L1b and L3b have somewhat poorer results. Model L3b has slightly better LLH/k scores than L2b,

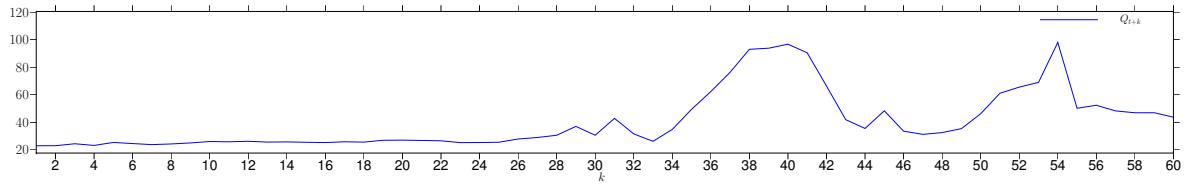
except for $k = 120$. With a couple exceptions, all three criteria indicate that Model 2a fits the data better than Model 1a, but both models are inferior to the three lag models. In general the MAD and RMSE scores are quite stable, and even improve from $k = 60$ to $k = 120$. As for the predictions made 30.08, the LLH/k scores are best for small k-values, due to increased uncertainty with k.



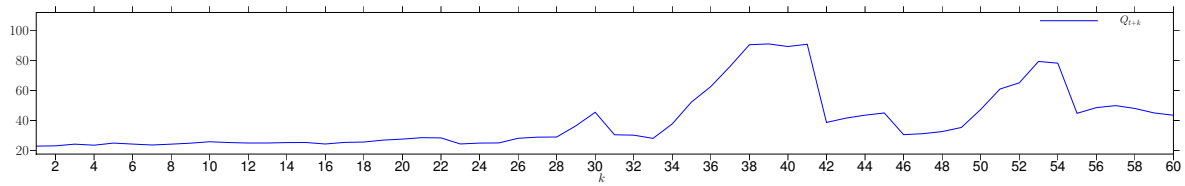
(a) M1a. k-values ranging from 1 to 12



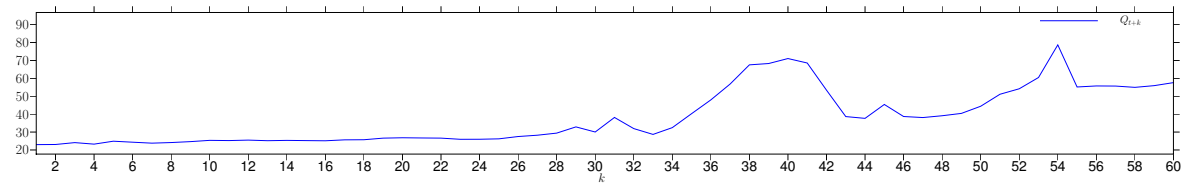
(b) M2a. k-values ranging from 1 to 36



(c) L1b. k-values ranging from 1 to 60



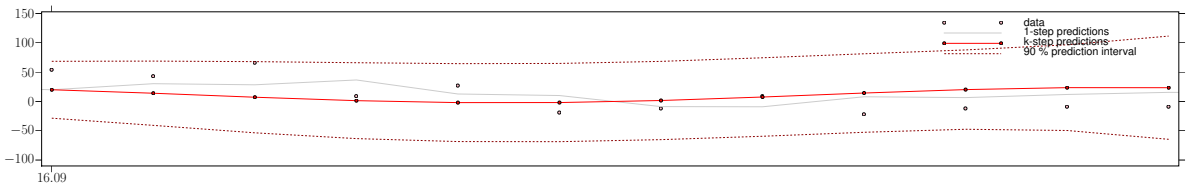
(d) L2b. k-values ranging from 1 to 60



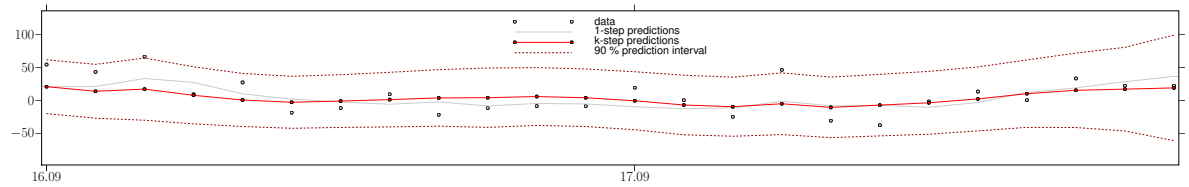
(e) L3b. k-values ranging from 1 to 60

Figure 6.5: Standard deviation for k-step predictions made 15.09 at 22.00 Hrs

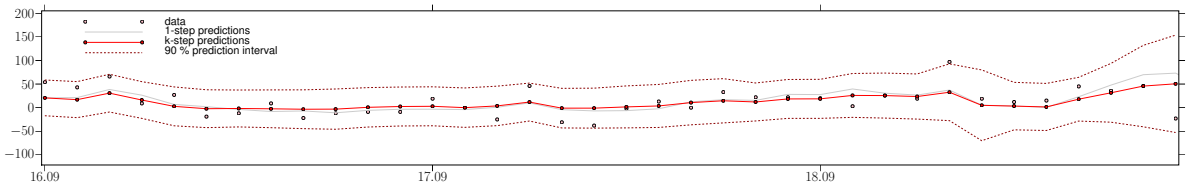
From Figure 6.5, we see once again that the prediction uncertainty increases rapidly for Model 1a. The k-step prediction standard deviation for Model 2a is quite stable till around $k = 20$, but increases rapidly from this point. For the lag models, the standard deviation is stable till around $k = 28$.



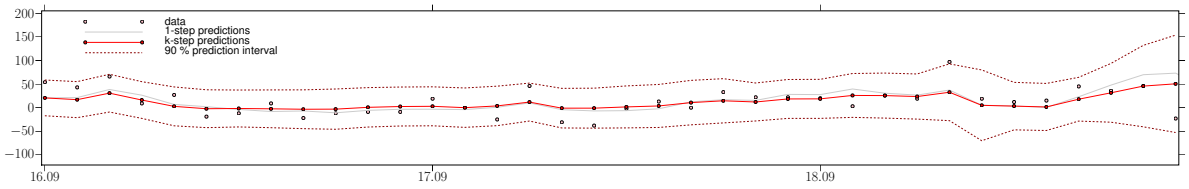
(a) M1a. k-values ranging from 1 to 12



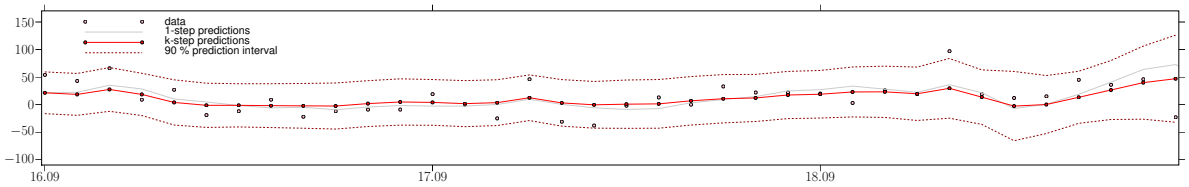
(b) M2a. k-values ranging from 1 to 24



(c) L1b. k-values ranging from 1 to 36



(d) L2b. k-values ranging from 1 to 36



(e) L3b. k-values ranging from 1 to 36

Figure 6.6: k-step predictions made 15.09 at 22.00 Hrs

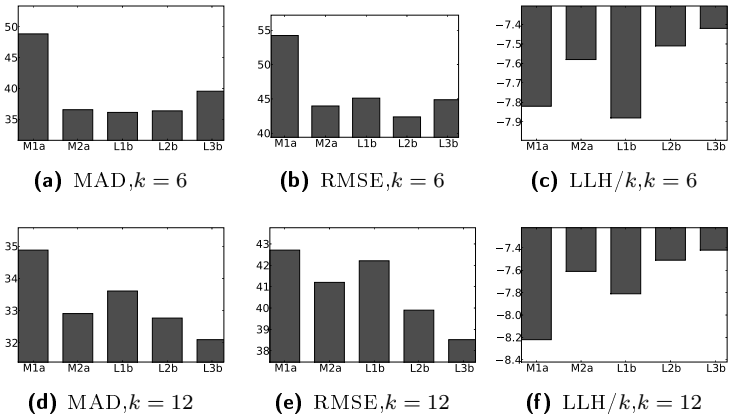
From Figure 6.6, we can see that the lag models have good point predictions and stable prediction intervals for more than two days. At the middle of the third day, around $k = 28$, the uncertainty increases, but most point predictions are still accurate. Only a few observations lie outside the 90 % prediction interval. Models 1a and 2a can not compete with the accuracy of the lag models, and the prediction intervals are not nearly as narrow, especially for Model 1a.

6.3. Results for Predictions made 01.10 at 22.00 Hrs

Performance scores for k-step predictions made 01.10 at 22.00 Hrs are given in Table 6.4 and shown in Figure 6.7. Model L3b is the best model for $k = 12$ according to all criteria, but has poor RMSE and MAD scores for $k = 6$. Model L1b has the best MAD score for $k = 6$, but

Model	k	MAD	RMSE	LLH/k
M1a	6	48.86	54.23	-7.82
M1a	12	34.88	42.71	-8.22
M2a	6	36.55	43.97	-7.58
M2a	12	32.91	41.20	-7.61
L1b	6	36.11	45.11	-7.88
L1b	12	33.61	42.21	-7.81
L2b	6	36.36	42.38	-7.51
L2b	12	32.77	39.90	-7.51
L3b	6	39.55	44.88	-7.42
L3b	12	32.10	38.52	-7.42

Table 6.4: Performance scores for k-step predictions made 01.10 at 22.00Hrs for Models 1a, 2a, L1b, L2b and L3b.



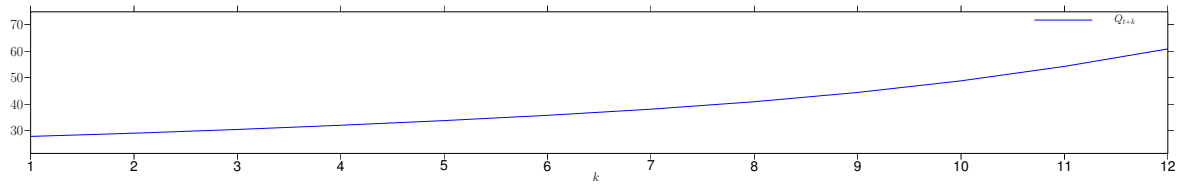
(a) MAD,$k = 6$ **(b)** RMSE,$k = 6$ **(c)** LLH/k,$k = 6$

(d) MAD,$k = 12$ **(e)** RMSE,$k = 12$ **(f)** LLH/k,$k = 12$

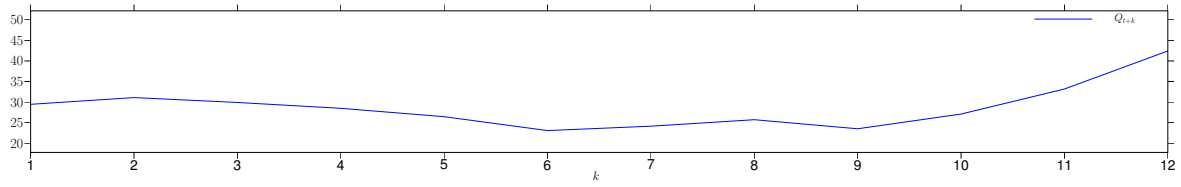
Figure 6.7: Performance scores for step ahead forecasting with Models 1a, 2a, L1b, L2b and L3b. k-values 6, 12, 24, 36, 48, 60 and 120 are used, which corresponds to respectively 1/2, 1, 2, 3, 4, 5 and 10 days. The predictions are made 01.10 at 22.00 Hrs

poor RMSE and LLH/k scores. For $k = 12$ Model 2a performs better than Model L1b. Even though Model L2b only has the best score for the RMSE for $k = 6$, it comes out as the winning model as it has an overall good performance compared to the other models. Model 1a again has the poorest prediction performance. Compared to the predictions made 30.08 and 15.09, the prediction performance is significantly worse. As there are only 12 data points available after 01.10 at 22.00 Hrs, we get a limited impression of the prediction performances.

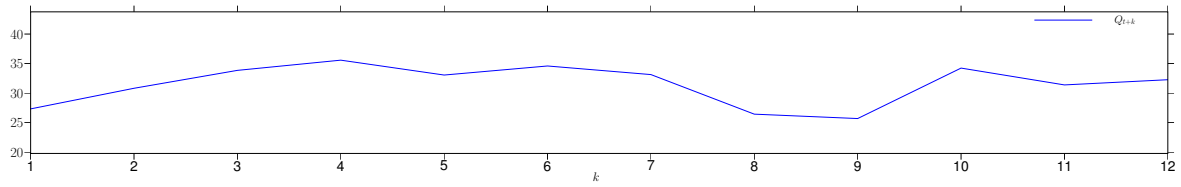
From Figure 6.8 we see that the standard deviation for the three lag models are quite stable till $k = 12$. For Models 1a and 2a the prediction uncertainty increases drastically within the 12 first steps. Figure 6.9 shows that 2 or 3 of the 12 observations lie outside the prediction intervals for all models. From Figure 3.4 (a) we can see that the strain data are quite volatile 02.10 and 03.10. Again this might imply that the models do a better job predicting more stable periods. In spite of some outliers, quite a few predictions are very accurate for both the lag models and Model 2a.



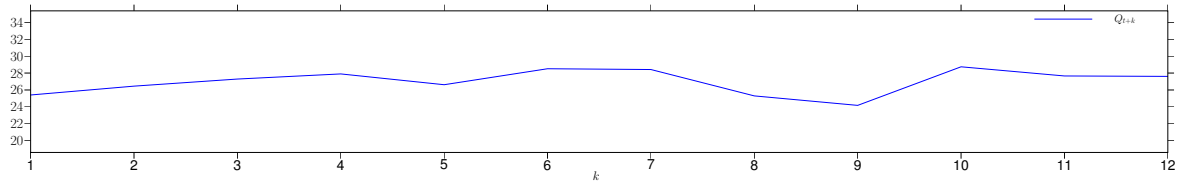
(a) M1a. k-values ranging from 1 to 12



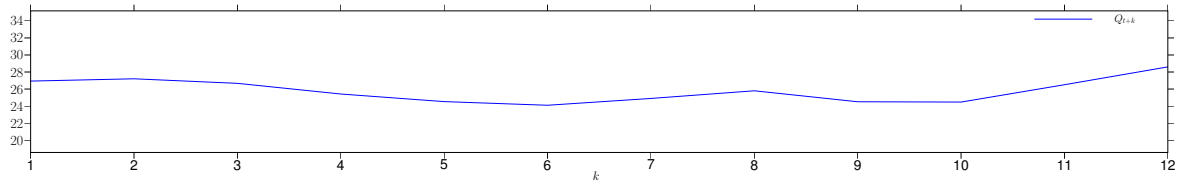
(b) M2a. k-values ranging from 1 to 12



(c) L1b. k-values ranging from 1 to 12

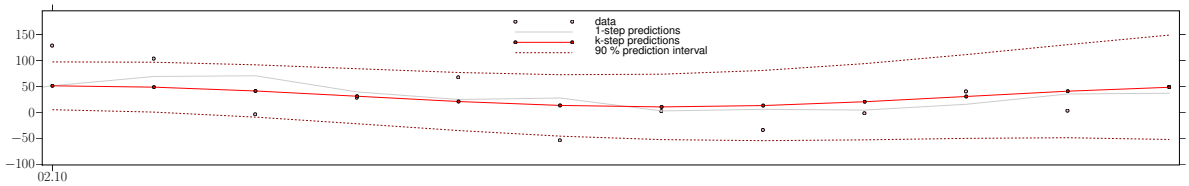


(d) L2b. k-values ranging from 1 to 12

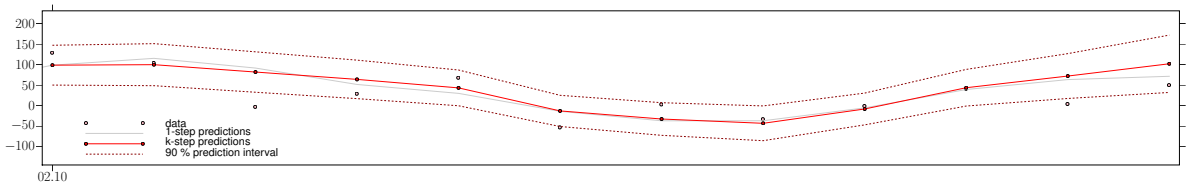


(e) L3b. k-values ranging from 1 to 12

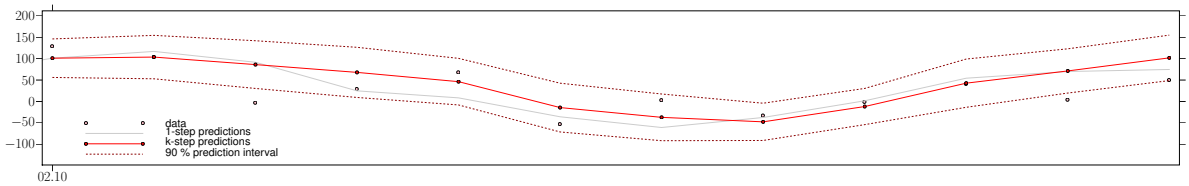
Figure 6.8: Standard deviation for k-step predictions made 01.10 at 22.00 Hrs



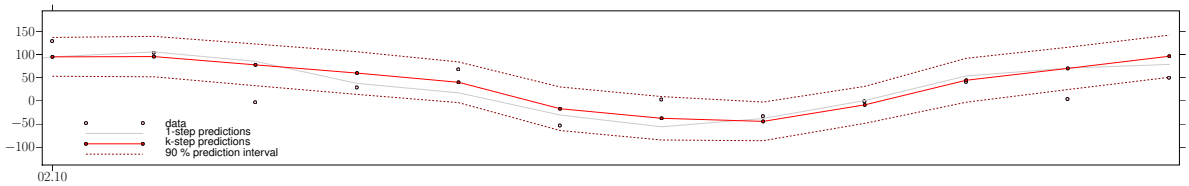
(a) M1a. k-values ranging from 1 to 12



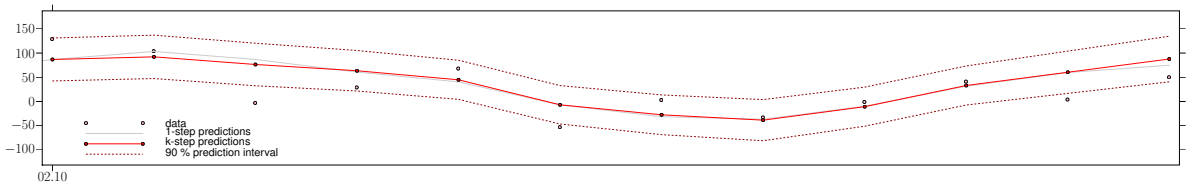
(b) M2a. k-values ranging from 1 to 12



(c) L1b. k-values ranging from 1 to 12



(d) L2b. k-values ranging from 1 to 12



(e) L3b. k-values ranging from 1 to 12

Figure 6.9: k-step predictions made 01.10 at 22.00 Hrs

6.4. Conclusions

Model L2b has the best performance, both for short and long term predictions. There are some exceptions, but all in all L2b produces good point predictions and also the prediction uncertainty is stable for about 2 to 3 days. There are indications that the model has better predictions for periods where the strain and temperature data are relatively stable. The point predictions can be of high quality for 10 days, as for the predictions made 15.09. However, the uncertainty of the predictions that are far into the future is large. If there are changes in parameters, point predictions many steps ahead might be of lower quality, as we experienced for predictions made 30.08. The increasing uncertainty reflects the possible deterioration of the predictions. A few observations lay outside the prediction intervals for Model L2b. We do not have any information on the quality or origin of these observations, so it is impossible to know if these are real outliers that should be discarded, or if the model is insufficient. There are no obvious differences in the outcome for the shifting methods, but the static shifted version of the lag models is the winning model. Even though the simpler regression model M2a had some good results, it can not compete with the lag models. The non-regression model M1a is inferior to both the lag models and to M2a.

The step ahead predictions are done based on known regression variables. In a real-time context future regressors will not be known. How to include uncertainty of regression variables into the model was discussed in Section 5.3.

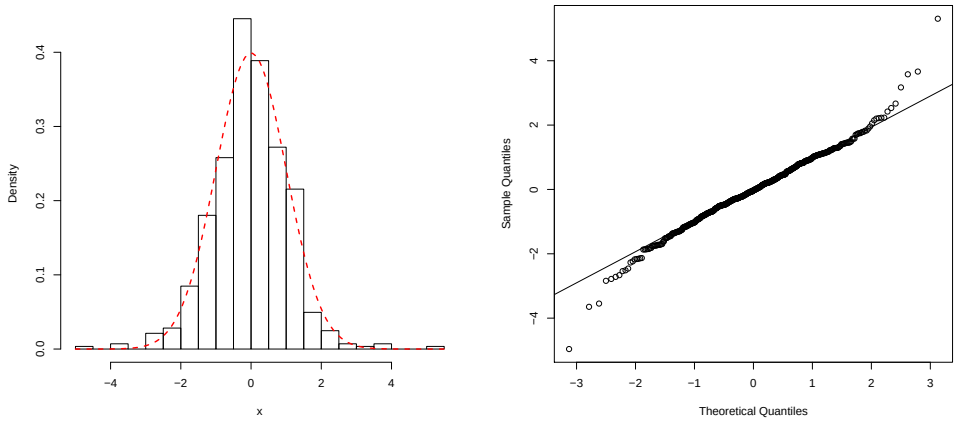
CHAPTER 7

Error Analysis

As residuals are the remainders that are left unexplained by the model, potential flaws and inadequacies in the model can be revealed by analyzing errors. In Chapter 5 and 6 we found that Model L2b had the best fit for the strain data. However, the model performance deteriorated for some short periods of time. In this chapter we will take a closer look at the residuals for Model L2b, and suggest solutions for improving the performance during these periods.

7.1. Normality Assumption

It is assumed that $e_t|D_{t-1} \sim T_{n_{t-1}}[0, Q_t]$. Hence, the standardized errors, $e_t/\sqrt{Q_t}|D_{t-1}$, should have a $T_{n_{t-1}}[0, 1]$ distribution. As we chose to have a 4 day burn-in period, corresponding to 48 observations, the error terms are T-distributed with 48 to 614 degrees of freedom. Thus we can assume approximately normality in the errors. As we see from Figure 7.1, the normality assumption is not violated. Both the histogram in Figure 7.1 (a) and the QQ-plot in Figure 7.1 (b) reveals somewhat heavy tails, but the standardized errors are still close to a Gaussian distribution.



(a) The histogram shows the standardized errors, $e_t/\sqrt{Q_t}$, for Model L2b. The dotted line is a standard normal distribution.

(b) Normal QQ-plot for $e_t/\sqrt{Q_t}$, for Model L2b

Figure 7.1: Error plots

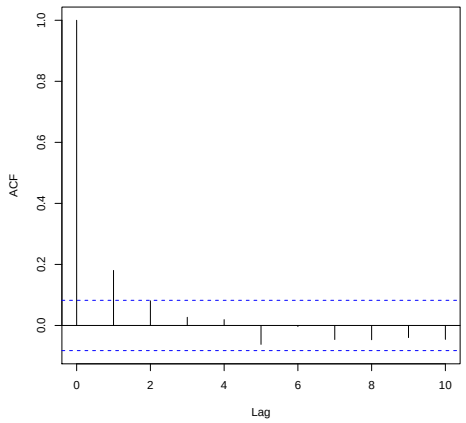


Figure 7.2: Autocorrelations for e_t with Model L2b

7.2. Sequential Correlations

Correlations in the errors were our main concern for Model L2b. The autocorrelation plot in Figure 7.2 confirms our suspicion, showing evidence of positive correlation for lag = 1. Studying Figure 5.11 (d) the correlations appear as occasional groups of sequential errors with the same sign. Several correlated errors as seen for Model L2b can indicate that the model is not performing adequately. There can be various underlying reasons, and there are several ways to deal with such behavior:

- Seasonal patterns in the time series which are not modelled can cause consecutive, positively correlated errors. This is not the case for Model L2b. First of all, there are no signs of seasonal movements in the errors. Secondly, Model L2a, which is equivalent to Model L2b with an additional season block, had poorer performance.
- Noise models, such as autoregressive models described in West & Harrison (1997) Section 9.4, can be used to model local, positive correlations in errors for short periods of time. Such models will often improve short term forecasting accuracy, but it is however not to be used to explain more global movements which can be modelled by a basic DLM.
- Negatively correlated errors can be caused by using a model that adapts too much to data. Removing unnecessary parameters in the DLM or increasing discount factors should solve the problem. The autocorrelation plot did not show any signs of such correlations.
- Positively correlated errors can be a sign of structural change in the time series. High discount factors, as found to be optimal for Model L2b, give much weight to prior beliefs. As a consequence it can take some time for the series to adapt to sudden changes. Intervention in the time series by lowering discount factors for a short period of time, will allow the parameters in the model to adjust more rapidly. In Chapter

5, the models with low trend discount factors had better 1-step predictions for certain extreme observations. In the succeeding chapter we will examine this alternative closer.

CHAPTER 8

Model Monitoring

As briefly discussed in Chapter 7, positively correlated errors as seen for Model L2b, can be a sign of structural changes in the time series. How much a parameter adapts to new data depends on the uncertainty of the parameter. The more uncertainty in the parameter, the more it will be influenced by new data. Discount factors determine how much variance is added to the parameter when forecasting. As for Model L2b, discount factors are normally set to a quite high value, indicating low information loss. As a consequence, it can take some time for the parameters to adapt to changes. By lowering discount factors for short periods of time, the parameters are allowed to adjust more rapidly. In this chapter we will discuss how to detect structural changes and how to perform necessary interventions when such changes occur.

8.1. Interventions

Dependent on the nature of change in the time series, different modes of interventions can be performed:

- Succeeding observations which deviate from the model forecast is an indication of a structural change. To reflect lack of confidence in the model parameters, the uncertainty of one or several of the parameters can be increased. This is done by temporarily lowering one or several discount factors, so that the parameters can adjust more rapidly.
- One extreme Y_t-value could be a potential outlier. A typical intervention for an extreme observation is to ignore Y_t and treat it as a missing value. As there is a possibility that the observation is not an outlier, but a sign of a structural change, uncertainty is added so that the parameters can adapt to potential changes.
- If there is any information available which implies that there will be a change in the model dynamics, the value of one or several parameters can be adjusted. This is rarely done without adding additional uncertainty.

The latter intervention mode would in practice demand subjective judgment. We do not know of any underlying causes in our application which would motivate level adjustments in the parameters, thus we will not use this mode. To determine when to intervene, a monitoring scheme must be implemented to signal any possible exceptions in the time series dynamics. This concept is called monitoring by exception. Such signals can be treated manually by a qualified person

who decides whether to intervene or not when a possible change is signalled. Since we do not have any specific information about the dynamics in the data or about possible outliers, we will implement an automatic monitoring scheme where pre-defined exceptions initiate different intervention modes.

8.2. Model Monitoring with Bayes Factor

Given a standard or routine model that can properly describe the past dynamics of the time series in question, monitoring is done by continuously checking the routine model's forecasting performance. If the time series does not change behavior, it's 1-step forecasts should agree well with the real observed values. If the behavior of the time series changes, then the routine model will not be adequate. To detect such abnormal behavior, the routine model is compared to alternative models which describe alternative scenarios. E.g, if the routine model has a constant trend and it is of interest to see if the level might change, than the routine model can be compared to an equivalent model, only with a shifted level. Several alternative models can be constructed to monitor for various scenarios. It is also a possibility to compare models with different structures, for example a constant trend model vs a linear trend model, but we will not discuss this any further in this thesis.

A monitoring system, which compare model performances, will induce a signal if any of the alternative models performs significantly better than the routine model. The monitoring scheme triggers a signal if one of these three exceptions occur:

(1) One single extreme observation that deviates significantly from the routine model.
(2) Several deviant observations that are not extreme enough to qualify as exception (1), but jointly deviates sufficiently from the routine model.
(3) Small deviations over time that are not large enough to qualify as exception (1) or (2), but persists for a long period of time.

Consider two different models, M_R and M_A, where M_R is the routine model and M_A is an alternative model. The Bayes factor based on the observed value Y_t is then defined as

$$H_t = \frac{p\left(Y_t|D_{t-1}, M_R\right)}{p\left(Y_t|D_{t-1}, M_A\right)}$$

The Bayes factor can in this context be interpreted as the relative predictive probability, or as the odds for M_R to M_A. If H_t is larger than 1, the observation Y_t is more likely to occur given the routine model, compared to the alternative model. If Bayes factor is smaller than 1, the observed Y_t is more probable given the alternative model.

As the Bayes factor does not take into account historic performance of the routine and alternative models, other measures must be included into the scheme to monitor for exceptions (2) and (3).

The Bayes factor for k consecutive observations $Y_t, Y_{t-1}, ..., Y_{t-k}$ is defined in West & Harrison (1997) Definition 11.1 as

$$H_t(k) = \prod_{r=t-k+1}^{t} H_r = \frac{p(Y_t, Y_{t-1}, ..., Y_{t-k+1}|D_{t-k}, M_R)}{p(Y_t, Y_{t-1}, ..., Y_{t-k+1}|D_{t-k}, M_A)}$$

The cumulative Bayes factor L_t is further defined in Theorem 11.3 as

$$L_t = \min_{1 \leq k \leq t} H_t(k),$$

where L_t is minimized at $k = l_t$ and

$$l_t = \begin{cases} 1 + l_{t-1}, & L_{t-1} < 1 \\ 1, & L_{t-1} \geq 1 \end{cases} \qquad t = 2, ...$$

If the routine model has performed better than the alternative model prior to time t, the cumulative Bayes factor L_t is merely the Bayes factor at time t. If the alternative model has performed better than the routine model, then L_t is the product of the preceding Bayes factors starting with the first Bayes factor that was in favor of the alternative model up till the current Bayes factor. Whenever the cumulative Bayes factor shows evidence in favor of the routine model, then L_t is reset to H_t, thus previous evidence is no longer taken into account. l_t is the run length of cumulative evidence in favor of the alternative model. If the alternative model has performed better than the routine model for some time, then l_t denotes the duration of this trend. If L_t is in favor of the routine model, then l_t is reset to 1. Thus the monitoring scheme can detect the 3 exceptions as follows:

- Exception (1): $H_t \leq \tau$
- Exception (2): $L_t \leq \tau$
- Exception (3): $l_t \geq l_{max}$,

where l_{max} is a threshold run length typically set to 3 or 4, and $\tau \in (0, 1)$ is a limit value for H_t and L_t. If τ is close to 1, the monitoring scheme will be very sensitive to even small deviations from the routine model, while a monitoring scheme with τ-value close to 0 would demand more evidence in favor of the alternative model to trigger a signal. Taking errors and chance into account, the following thumb rules for Bayes factor are used by Pole et al. (1994):

If $H_t > 10$: indicates evidence in favor of M_R

If $H_t > 100$: strong evidence in favor of M_R

If $H_t < 1/10$: indicates evidence in favor of M_A

If $H_t < 1/100$: strong evidence in favor of M_A

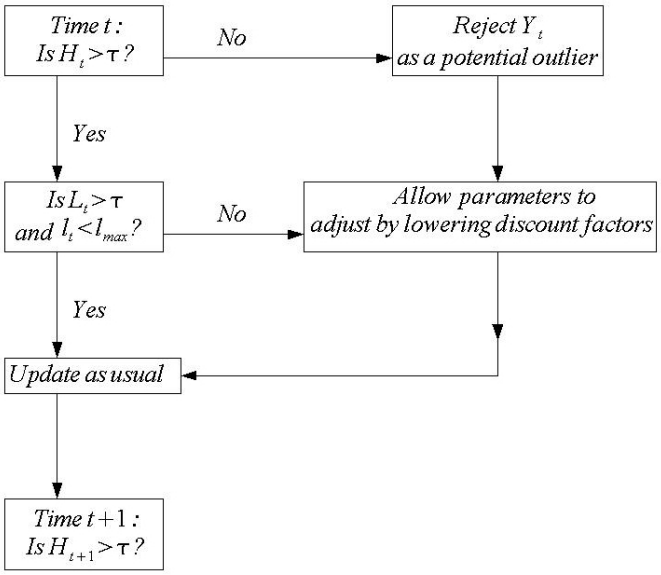


Figure 8.1: Automatic monitoring scheme by West & Harrison (1997) for detecting exceptions and making interventions.

If exception (1) is signalled, the automatic monitoring scheme will reject Y_t as an outlier and treat it as a missing value. In addition the parameter uncertainties will be increased, to let the parameters adjust in case of a structural change. If exceptions (2) or (3) occur, the parameter uncertainties are increased so the parameters can adjust to a new level. In all three cases the uncertainty is increased by using an alternative discount factor, denoted δ^{alt}. The monitoring scheme is summarized in Figure 8.1.

8.3. Monitoring for Level-Shifts

In Section 11.4.3 in West & Harrison (1997), monitoring for level-shifts is discussed for normally distributed errors. In this chapter we will have a parallel discussion for T-distributed errors. According to our assumption about the routine model M_0, e_t follows a $T_{n_{t-1}}[0, Q_t]$ distribution, and thus $e_t/\sqrt{Q_t} \sim T_{n_{t-1}}[0, 1]$. Let M_A be an alternative model equal to M_0, but with a level shift, so that e_t has a non-zero mean h. Then the Bayes factor at time t is defined as

$$H_t = \frac{p\left(e_t | D_{t-1}, M_R\right)}{p\left(e_t | D_{t-1}, M_A\right)}$$

Furthermore, we have that

$$p\left(\frac{e_t}{\sqrt{Q_t}} | D_{t-1}, M_0\right) = \frac{\Gamma\left(\left(n_{t-1}+1\right)/2\right)}{\sqrt{n_{t-1}\pi}\Gamma\left(n_{t-1}/2\right)} \left(1 + \frac{\left(e_t/\sqrt{Q_t}\right)^2}{n_{t-1}}\right)^{-(n_{t-1}+1)/2}$$

$$p\left(\frac{e_t}{\sqrt{Q_t}} | D_{t-1}, M_A\right) = p\left(\frac{e_t - h}{\sqrt{Q_t}} | D_{t-1}, M_0\right) = \frac{\Gamma\left(\left(n_{t-1}+1\right)/2\right)}{\sqrt{n_{t-1}\pi}\Gamma\left(n_{t-1}/2\right)} \left(1 + \frac{\left(\left(e_t - h\right)^2/\sqrt{Q_t}\right)^2}{n_{t-1}}\right)^{-(n_{t-1}+1)/2}$$

Thus

$$H_t = \frac{p\left(\frac{e_t}{\sqrt{Q_t}}|D_{t-1}, M_0\right)}{p\left(\frac{e_t}{\sqrt{Q_t}}|D_{t-1}, M_A\right)} = \frac{\left(1 + \frac{\left(e_t/\sqrt{Q_t}\right)^2}{n_{t-1}}\right)^{-(n_{t-1}+1)/2}}{\left(1 + \frac{\left((e_t-h)/\sqrt{Q_t}\right)^2}{n_{t-1}}\right)^{-(n_{t-1}+1)/2}}$$
$$= \left(\frac{Q_t n_{t-1} + {e_t}^2 - 2he_t + h^2}{Q_t n_{t-1} + {e_t}^2}\right)^{(n_{t-1}+1)/2}$$

If $H_t = 1$, then Bayes factor does not favor any of the models and

$$\left(\frac{Q_t n_{t-1} + {e_t}^2 - 2he_t + h^2}{Q_t n_{t-1} + {e_t}^2}\right)^{(n_{t-1}+1)/2} = 1$$
$$h(2e_t - h) = 0 \tag{8.1}$$

If $h = 2e_t$, then the Bayes factor does not favor any of the models. This seems reasonable as the observed error e_t is equally distanced to 0 and h, corresponding to the expected errors for the routine and the alternative model respectively. Based on the above discussion a natural choice for h at time t is

$$h_t = \pm 2t_\alpha \sqrt{Q_t},$$

where t_α is the α quantile in the $T_{n_{t-1}}$-distribution. If $H_t = 1$, indicating indifference between the two models, then according to (8.1)

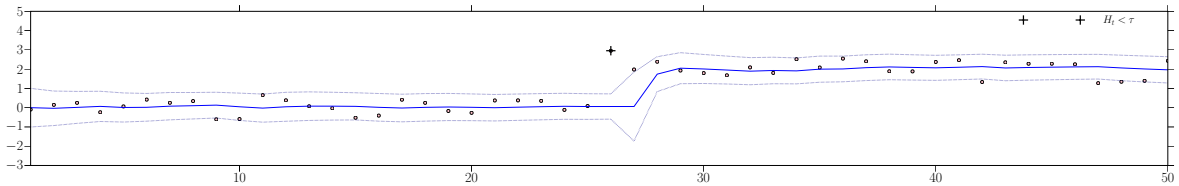
$$e_t = \frac{h_t}{2} = \pm t_\alpha \sqrt{Q_t}$$

Thus if e_t is equal to an acceptable upper or lower quantile of the error distribution $T_{n_{t-1}}[0, Q_t]$, the Bayes factor does not favor any of the models. Let us assume that α is chosen to be 0.95. Then $[-t_\alpha\sqrt{Q_t}, +t_\alpha\sqrt{Q_t}]$ forms a 90% prediction interval for e_t under the routine model. Using the suggested h_t, the Bayes factor will be in favor of the alternative model if the observed e_t falls outside the prediction interval. If e_t on the other hand is included in the interval, the Bayes factor will be larger than 1, and thus in favor of the routine model.

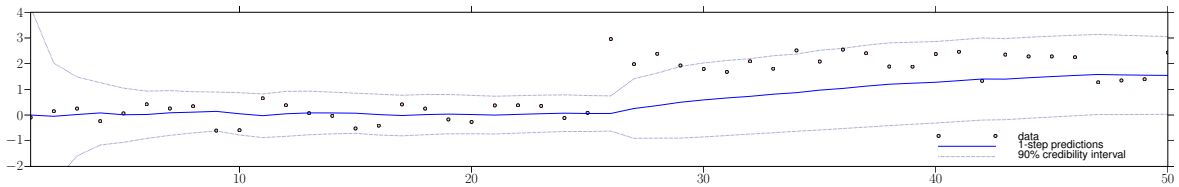
τ determines how strong the evidence against the routine model must be to induce an intervention. $H_t < 1$ favors the alternative model, but to trigger a signal, H_t must be below or equal to the threshold τ. In Example 11.2 in West & Harrison (1997) the choice of τ is discussed for standard normal distributed noise. They suggest to decide on an appropriate value for e_t that should imply indifference between the two models, and another value that should induce an intervention based on H_t. According to the limits set for acceptable errors, proper values for h and τ are chosen. However, the corresponding discussion for $T_{n_{t-1}}$-distributed noise with non-constant variance is much more complex. A possibility is to extend the discussion in West & Harrison (1997) to a time dependent τ. In this thesis we will however use a constant τ value. We will try out different τ-values, partly based on the previously mentioned thumb rules used in Pole et al. (1994).

Together with τ, the choice of α determines the sensibility of the monitoring scheme. While h, which depends on α, decides how extreme an observation must be for the Bayes factor to be in favor of the alternative model, τ determines how strong the evidence must be in favor of M_A to induce an intervention. In the following synthetic case, an automatic monitoring scheme which detects outliers and level jumps is implemented. Note that since we are monitoring for a level-shift, and not shifts in other parameters, the alternative discount factor δ_T^{alt} only affects the uncertainty of the level parameter.

EXAMPLE 8.1. *The synthetic data constructed for this example is normally distributed with a zero mean and* 0.2 *variance up till time* $t = 25$. *At* $t = 26$ *the mean level shifts to* 2. *In Figure 8.2(b) a 1st order polynomial trend model is applied. This is a proper choice of model, but without any intervention it takes a long time for the model to adjust to the correct level. Note that a discount factor 0.95 has been used, and that a lower discount factor would allow the level to adjust more rapidly. In Figure 8.2 (a) the automatic monitoring scheme is applied, and the level adjusts much faster. When the first exception is signalled at time* $t = 26$, *the monitoring scheme recognizes it as a potential outlier, deletes it from the data set and treats it as an outlier. In addition, the uncertainty is increased in case of real change. When* Y_{27} *is observed at the same level as* Y_{26}, *the level parameter quickly adjusts. (Note that if* Y_{27} *had signalled an exception as well, the observation would be discarded as an outlier and the level shift would have been postponed.) After a few observations, the uncertainty is back to normal. An alternative approach is to have a second scheme where not only the uncertainty is increased, but the expected level is adjusted as well. The level would then converge faster. Other deciding factors are the sensitivity of the monitoring scheme, and the increase of uncertainty, set by* δ_T^{alt}.



(a) With automatic interventions. $\tau = 0.10$, $\alpha = 0.95$, $\delta_T^{alt} = 0.05$ and $l_{max} = 3$



(b) Without interventions.

Figure 8.2: *1-step predictions with 90% prediction interval.*

It is always optimal to have a qualified decision maker to respond to alarms or warnings triggered by the monitoring scheme, instead of having automatic responses for pre-specified behavior, at

least to decline or allow for these interventions. This allows the decision maker to utilize any insight and information she or he may have concerning the data.□

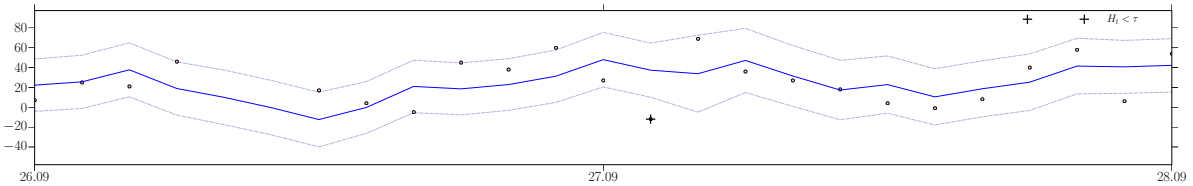
8.4. Monitoring level shifts in Model L2b

Before testing how the automatic level-shift monitoring scheme works for the bridge data, it is interesting to take a look at Figure 5.11 (b) which shows the posterior level parameter $\alpha_t|D_t$ for Model L2b. We notice several drops in the level at 23.08, 27.08 (temporary), 21.09 and 29.09. Hopefully the monitoring scheme will detect these exceptions, so the level adjusts more rapidly and the errors, which are particularly large around 27.08 and 29.08, are reduced.

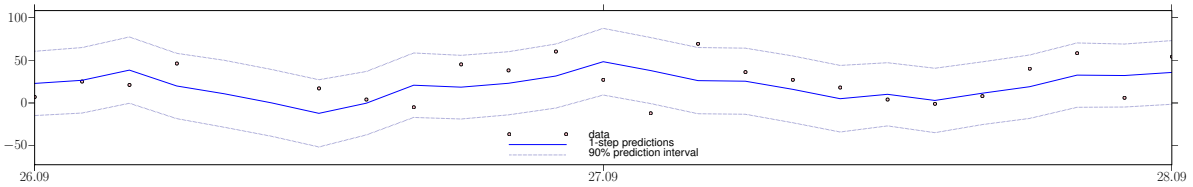
The sensitivity of the monitoring scheme, and the degree of intervention is decided by the parameters τ, t_α and δ_T^{alt}. Testing out alternative values, the automatic monitoring scheme shows several severe flaws. If the Bayes factor is less than or equal to the set threshold τ, the scheme detects an exception which calls for intervention. Y_t is treated as a missing value, and the uncertainty around the level parameter is increased to let the parameter adjust to a possible structural change. After this type of intervention, one of two scenarios is accounted for. If Y_t is a real outlier, Y_{t+1} is expected to follow the same dynamics as prior to the intervention. If this is the case, the level parameter will stay more or less unchanged, and the extra added uncertainty will be reduced. In the second scenario, Y_t is not an outlier, but the first observation after a level shift. Thus Y_{t+1} will follow at the same level as Y_t, just as in the previous example, and the level parameter will adjust quickly (given that Y_{t+1} does not signal an exception). A third scenario which have not been accounted for, occurs if the level in Y_{t+1} does not coincide with neither the Y_t-level, nor the level prior to the intervention. In the following example we will take a closer look on how an intervention can affect the predictions if this is the case.

EXAMPLE 8.2. *In this example we will apply the automatic level-shift monitoring scheme to the bridge data. We set* $\tau = 0.1$, $t_\alpha = t_{0.95}$, $\delta_T^{alt} = 0.05$ *and* $l_{max} = 3$. *As we see from Figure 8.3 (a),* Y_{565} *observed 27.09 at 02.00Hrs signals an exception since* $H_t \leq \tau$. *As the uncertainty is considerably increased by the intervention, the succeeding observation* Y_{566} *will as a consequence have great influence on the level parameter. The intention of increasing the uncertainty is to allow the level to adjust to a possible lowered level, as the observation* Y_{565} *may represent.* Y_{566} *does however not have a lowered level, but to the contrary it lies quite high above the 1-step prediction. Consequently the level is adjusted upwards because of the initial suspicion of a downwards level shift. Comparing the predictions to the non-monitored results in Figure 8.3(b), the latter is not nearly as much influenced by the observation* Y_{566}. *This shows that it is questionable to uncritically increase the parameter uncertainty. Especially if the alternative discount factor is low, the level parameter becomes very vulnerable subsequent to an intervention.*

□



(a) With automatic interventions. $\tau = 0.1$, $\alpha = 0.95$, $\delta_{Talt} = 0.05$, $l_{max} = 3$.



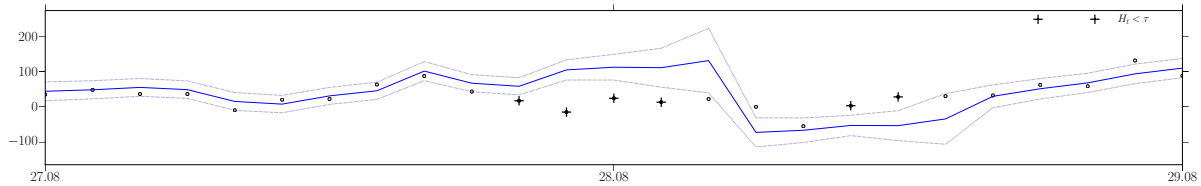
(b) Without interventions.

Figure 8.3: *1-step predictions with a 90 % prediction interval.*

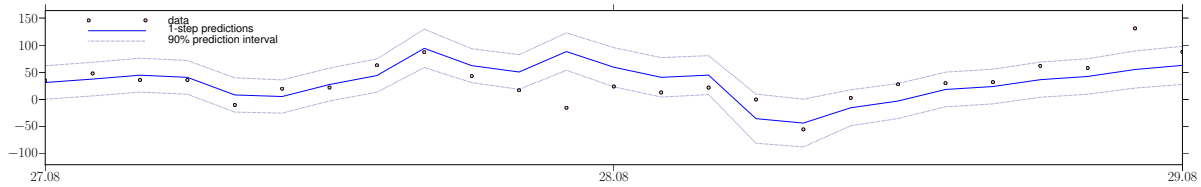
When using relatively high alternative discount factors, another serious flaw with the automatic intervention scheme is revealed. When an observation is deleted from the data set due to the exception $H_t \leq \tau$, it is not taken into account that the following observation might also induce the same type of exception. In that case, we run the risk of dismissing several succeeding observations as outliers. In the following example we illustrate how this can potentially lead to overlooking valuable information.

EXAMPLE 8.3. *In this example we apply the automatic monitoring scheme to the bridge data, with* $\tau = 0.1$, $t_\alpha = 0.95$, $\delta_T^{alt} = 0.3$ *and* $l_{max} = 3$. *The results in Figure 8.4 (a), show 4 succeeding exceptions starting 27.08 at 20.00Hrs where* $H_t \leq \tau$ *for all four observations* Y_{202}, Y_{203}, Y_{204} *and* Y_{205}. *As a consequence, the observations are all deleted from the dataset, and the level parameter is not adjusted at all before observing* Y_{206} *28.08 at 04.00Hrs, the first observation that does not trigger an intervention. Since there have been four succeeding interventions, the forecast uncertainty has grown large. Comparing the predictions with the non-monitored version in Figure 8.4 (b), wee see that the latter looks more reasonable. By performing the interventions we ignore four succeeding observations that are not outliers. Because of the added uncertainty due to the interventions, the level parameter adjusts too much, and as a consequence a new signal is triggered by* Y_{209}, *and another two observations* Y_{209} *and* Y_{210} *are dismissed from the dataset.* □

Despite the drawbacks, we continue with finding the τ, t_α and δ_T^{alt} values which give the best results according to RMSE, MAD and LLH/T for the strain data. However, we avoid deleting potential outliers, but merely increase the uncertainty if $H_t \leq \tau$. The best results are achieved with $\tau = 0.01$, $t_\alpha = t_{0.995}$ and $\delta_T^{alt} = 0.25$. l_{max} is set to 3. The RMSE score is 23.41, the MAD score 17.75 and LLH/T is -6.96, a slight improvement for the RMSE and LLH/T which originally scored respectively 23.44 and -6.97 for Model L2b. The MAD was earlier 17.70, thus



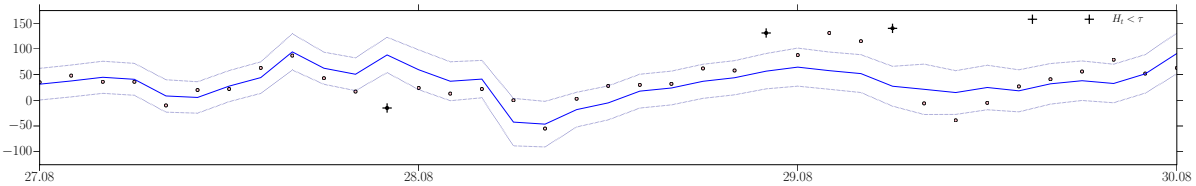
(a) With automatic interventions. $\tau = 0.1$, $\alpha = 0.95$, $\delta_T{}^{alt} = 0.3$, $l_{max} = 3$.



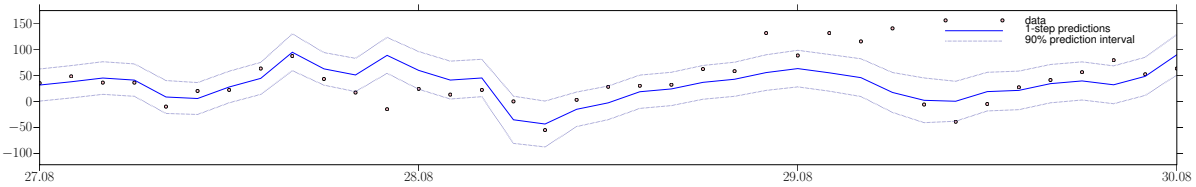
(b) Without interventions.

Figure 8.4: *1-step predictions with a 90 % prediction interval*

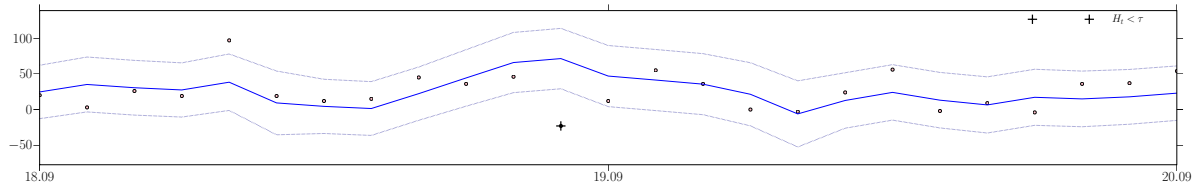
the monitored version performs poorer according to MAD. The chosen τ and t_α values give a monitoring scheme with very low sensitivity. h is set to twice the upper and lower 0.995-quantile of the forecast distribution and $\tau = 0.01$ demands that the evidence in favor of the alternative model must be very strong to induce any interventions. With the given τ, t_α, δ_T^{alt} and l_{max} values, four automatic interventions are made throughout the dataset, all due to H_t exceeding the threshold τ. The three first interventions, triggered by Y_{203}, Y_{215} and Y_{219}, in the time span between 27.08 and 29.08, are shown in Figure 8.5 (a). The corresponding 1-step predictions without interventions are shown in Figure 8.5 (b). The fourth interventions is signaled by Y_{467}, 18.09 at 22.00Hrs. Figure 8.5 (c) shows the 1-step predictions from 18.09 to 20.09 with interventions, while Figure 8.5 (d) shows the corresponding predictions without interventions. The differences between the monitored and non-monitored versions are minimal. All in all the monitoring gives disappointing results, and reveals serious flaws in the scheme. Handling monitoring signals manually reduces the risk of performing incorrect interventions. If a situation occurs which implies higher uncertainties around the parameters, e.g. very heavy traffic load or extreme temperatures, interventions can be done manually. For the dataset analyzed we have no such information, and the automatic monitoring scheme does not significantly improve the forecast performance.



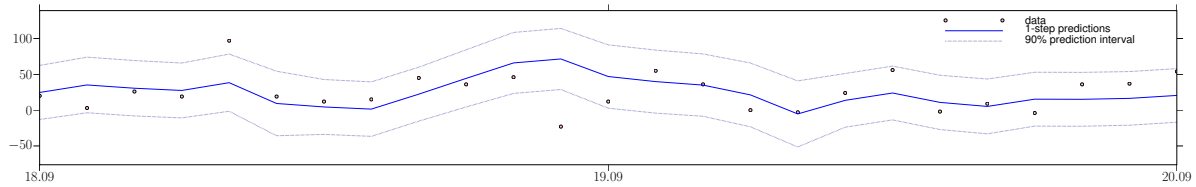
(a) With automatic interventions. $\tau = 0.01$, $\alpha = 0.995$, $\delta_{Talt} = 0.25$, $l_{max} = 3$.



(b) Without interventions



(c) With automatic interventions. $\tau = 0.01$, $\alpha = 0.995$, $\delta_{Talt} = 0.25$, $l_{max} = 3$.



(d) Without interventions

Figure 8.5: 1-step predictions with a 90 % prediction interval. (a), (b): 27.08-30.08. (c), (d): 18.09-20.09.

CHAPTER 9

Conclusions and discussions

The main objective of this thesis was to find a time series model which produces good 1-step and several step ahead forecasts for real-time strain data from the Göta River Bridge, using the Bayesian time series framework described in West & Harrison (1997). As temperature affects strain, we introduced air temperature and radiation as possible explanatory variables. Based on the predictive performance measures RMSE, MAD and LLH, we evaluated different trend, season and regression models. We concluded that a form-free transfer function which regresses on current and previous air temperatures has the overall best performance. To avoid that the regression variables influence the level parameter, the variables had to be shifted so that the average contribution from the regressor is equal to zero. We introduced two different shifting alternatives. Dynamic shifting gave a level parameter which reflected the real underlying level of the data, while using static shifting gave the best performance and step ahead prediction results. Having in mind that the original accuracy of the strain measurement is plus/minus 21 micro strain, the latter model, Model L2b, achieved good results. We also partly achieved very good long term predictions. However, the model gives better predictions for stable data, while the performance deteriorates when the data are very unstable. The more unstable periods seem to occur when there are large variations in temperatures during a day, and our model is somewhat conservative when this occurs.

As the model includes air temperatures, forecasting requires regressors which would not be known in a real-time setting. We can solve this problem either by introducing the uncertainty of the regressors into the model, or by doing 'What if?' predictions for given scenarios. In our application we focused on finding the best model for the data, thus we used information which in practice would demand foresight.

The normal assumption for the errors is not violated, but we found evidence of positive correlations for subsequent errors. One possible explanation for such behavior can be high discount factors, which do not allow rapid changes in the model parameters. Hence we suggested temporarily lowering discount factors where there were evidence of change in the parameters. We used the monitoring scheme based on Bayes factor, and with automatic interventions as described in West & Harrison (1997). We discovered that the automatic interventions suggested, are not very well considered. If we use the monitoring and intervention scheme uncritically, we run the risk of incorrectly making adjustments in the parameter values, and dismissing valuable data as

outliers. West & Harrison (1997) suggest another change detection scheme based on cumulative sums, which takes loss functions and decision theory into consideration in Section 11.6.4.

By dividing the dataset into three parts, we discovered that information loss (expressed through discount factors) is not constant, but varies over time. We tried lowering discount factors temporarily, but this was primarily motivated by potential abrupt level changes in parameter values. A more sophisticated procedure to allow variations in discount factors is to use multi-process models, class I or II, as described in Chapter 12 in (West & Harrison, 1997). For *multi-process models, class I*, one specific DLM is assumed to be the correct model for all times t, but some parameters, $\boldsymbol{\alpha}$ are subject to uncertainty. α can represent any defining parameter in the model. For this specific purpose $\boldsymbol{\alpha}$ would represent the vector of discount factors. Let $\mathcal{A}$ denote possible sets of values for $\boldsymbol{\alpha}$, and let $\{M_t(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \mathcal{A}\}$ be the class of possible DLMs at time t. The 'true' value of $\boldsymbol{\alpha}$ is unknown, but for some $\boldsymbol{\alpha}_0 \in \mathcal{A}$, $M_t(\boldsymbol{\alpha}_0)$ holds for all time t. Using *multi-process models class II*, it is not assumed that one DLM is the correct model for all times t, but that different models are appropriate for different times. Thus for some sequence of values $\boldsymbol{\alpha}_t \in \mathcal{A}$, $M_t(\boldsymbol{\alpha}_t)$ holds for time t, $(t = 1, 2, ...)$. According to West & Harrison (1997), the latter approach is more realistic, and also the natural choice if we assume that discount factors actually do change over time (in contrast to there being one unknown 'true' combination of discount factors). Given a single block model with one discount factor, let $\alpha = \delta$ with a set of possible values $\mathcal{A} = \{0.9, 1.0\}$. Then the discount factor δ_t takes either the value 0.9 or the value 1.0. By using multi-process, class II *mixture* models, the predictive density is a linear combination of the alternative models. Individual models are weighted due to their respective probability for being the 'true' model at any time. Thus with $\mathcal{A} = \{0.9, 1.0\}$, δ_t can take any value between 0.9 and 1.0. Using the multi-process, class II mixture model with a set of different discount factors $\mathcal{A}$, would be a natural step to further improve our model. Some restrictions and assumptions must be done to limit computational demands to an acceptable level.

Allowing dynamic discount factors would probably improve our model, but not solve the problem regarding positively correlated errors. In Chapter 7 we briefly discussed the idea of introducing noise models for short periods of time to model local correlation structures as experienced for Model L2b. A DLM with an included autoregressive term is defined by

$$Y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + X_t + \nu_t, \qquad \nu_t \sim T_{n_{t-1}}[0, V] \tag{9.1}$$

where X_t is an added unobservable quantity describing behavior which is not adequately explained by the mean response function $\mathbf{F}_t'\boldsymbol{\theta}_t$. A way of introducing an autoregressive term for restricted periods of time, is to use a multi-process, class II model, with a DLM as defined in (9.1). Furthermore, let α_t represent X_t, $\alpha_t = X_t$. X_t can take on the set of values $\mathcal{A} = \{0, \tilde{X}_t\}$, where $\tilde{X}_t$ is defined by some noise process term. If $\alpha_t = X_t = 0$, then the autoregressive term would have no influence on the DLM. Section 9.4 in West & Harrison (1997) introduces a range of different noise models, which can be explored to improve our model.

As there are over 50.000 monitored points on the bridge, the univariate time series could be extended to include several points, using multivariate models as described in West & Harrison (1997) Chapter 16. In this thesis, we have focused on constructing a good model for a univariate strain time series, as this lay the foundation for a multivariate series.

Bibliography

Benham, P., Crawford, R. & Armstrong, C. (1996). *Mechanics of Engineering Materials. Second Edition.* Longman.

Berger, J. (1998). *Statistical Decision Theory and Bayesian Analysis. Second Edition.* Springer.

Glišić, B., Posenato, D. & Inaudi, D. (2007). Integrity monitoring of old steel bridge using fiber optic distributed sensors based on Brillouin scattering. *14th SPIE Annual Symposium on Smart Structures and Materials and Nondestructive Evaluation and Health Monitoring* **6531-25**.

Langtangen, H. (2008). *Python Scripting for Computational Science. Third Edition.* Springer.

Myhre, E. (1997). *Modellering ved Bayesiansk Tidsrekkeanalyse på Motorvognskader.* Universitetet i Oslo.

Natvig, B. (1997). *En Introduksjon til Bayesiansk Statistikk og Beslutningsteori. 2. utgave.* Matematisk institutt, Unversitetet i Oslo.

Natvig, B. & Tvete, I. F. (2007). Bayesian hierarchical space-time modeling of earthquake data. *Methodology and Computing in Applied Probability* **9**, 89–114.

NGI (2006a). Clarification of specifications Matpres 108.1, 109.1, 111.1.

NGI (2006b). Definition of the bridge coordinate system.

NGI (2008a). Optical fibres safeguard bridge. Available at: `http://www.ngi.no/en/Contentboxes-and-structures/Reference-Projects/Reference-projects/The-Gotariver-Bridge-safe-with-optical-fibres/`.

NGI (2008b). Temperature measurements related to girders for the period 10 december 2007 - 20 may 2008.

Pole, A., West, M. & Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis.* Chapman & Hall.

Spall, J. (1988). *Bayesian Analysis of Time Series and Dynamic Models.* Marcell Dekker, Inc.

Tent, M. (2006). *The Prince of Mathematics: Carl Friedrich Gauss.* A K Peters, Ltd.

Tvete, I. (2000). *Bayesiansk tidsrekkeanalyse anvendt på næringslivsforsikringer.* Universitetet i Oslo.

Tvete, I. & Natvig, B. (2002). A comparison of an analytical approach and a standard simulation approach in Bayesian forecasting applied to monthly data from insurance companies. *Methodology and Computing in Applied Probability* **4**, 95–113.

West, M. (1997). Bayesian forecasting. In *Encyclopedia of Statistical Sciences*, S. Kotz, C. Read & D. Banks, eds. Wiley.

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer.